ESIG
Large Loads
Task Force

Large Loads

BEHAVIORS, CAPABILITIES, AND LIMITATIONS

A Report by the
Energy Systems Integration Group's
Large Loads Task Force

July 2026

ESIG
ENERGY SYSTEMS
INTEGRATION GROUP

About ESIG

The Energy Systems Integration Group is a nonprofit organization that marshals the expertise of the electricity industry's technical community to support grid transformation and energy systems integration and operation. More information is available at https://www.esig.energy.

ESIG Publications Available Online

This report is available at https://www.esig.energy/working-groups/large-loads/. All ESIG publications can be found at https://www.esig.energy/resources/reports-and-briefs/.

Get in Touch

To learn more about the topics discussed in this report or for more information about the Energy Systems Integration Group, please send an email to info@esig.energy.

Large Loads: Behaviors, Capabilities, and Limitations

A Report by the Energy Systems Integration Group's Large Loads Task Force

Prepared By

Eric Meier, Electric Reliability Council of Texas

Jarrad Wright, National Laboratory of the Rockies

Sergey Kynev, Siemens Energy

Amin Dadashzade, Zero-Emission Grid

Omid Alizadeh, Danovo

Nihal Mohan, NextEra Energy Transmission

Novan Zakkia, Hitachi Energy

Inalvis Alvarez Fernandez, Simple Thread

Project Manager

Julia Matevosyan, Energy Systems Integration Group

Report Contributors

Ahmed Rashwan, Electric Power Engineers

Nicholas Zagrodnik, WEC Energy Group

Evan Neel, Lancium

Preface to ESIG Large Loads Task Force Reports

This report is one of 11 reports by the ESIG Large Loads Task Force, which was formed to assist the power industry in addressing new challenges introduced by the rapid proliferation of large electronic loads such as data centers, as well as other large loads including manufacturing, electric vehicle fleets, and hydrogen production. The titles of the reports are as follows:

- Grid Integration of Large Loads: Introduction to the Large Loads Task Force, Data Needs, and Flexibility
- Forecasting for Large Loads: Current Practices and Recommendations
- Interconnection Processes for Large Loads: Current Practices and Recommendations
- Large Load Performance Requirements: Current Practices and Recommendations
- Large Loads: Behaviors, Capabilities, and Limitations
- Reliability Impacts of Large Loads
- Large Load Disturbance Events
- Large Load Modeling for Dynamic Studies: Current Practices and Recommendations
- Transmission Planning with Large Loads: Current Practices and Recommendations
- Resource Adequacy with Large Loads: Planning for Flexibility to Accelerate Integration
- Wholesale Market Design and Operations for Systems with Large Loads: Current Practices and Recommendations

Disclaimer

Suggested Citation

Energy Systems Integration Group. 2026. *Large Loads: Behaviors, Capabilities, and Limitations.* A report by the Large Loads Task Force. https://www.esig.energy/reports-briefs/large-loads-behaviors-capabilities.

Contents

Introduction 1
Data Centers 4
Facility Design 4
Operational Characteristics 20
Cryptocurrency Mining 30
Facility Design 30
Operational Characteristics 34
Electrical Performance 36
Hydrogen Production 40
Facility Design 41
Operational Characteristics 46
Electrical Performance 50
Appendix: Overview of Large Load Categories and Configurations 53
Category 1: Data Centers and Computational Load 53
Category 2: Industrial Load Types 58
Category 3: Hydrogen Production Facilities 63
Category 4: Transportation Load 64
References 66

Abbreviations

AEL	Alkaline electrolyzer
AEM	Anion exchange membrane [electrolyzer]
AI	Artificial intelligence
ASIC	Application-specific integrated circuit
BBU	Battery backup unit
CPU	Central processing unit
DDL	Distributed deep learning
EMT	Electromagnetic transient
ERCOT	Electric Reliability Council of Texas
EV	Electric vehicle
GPU	Graphics processing unit
IGBT	Insulated gate bipolar transistor
IT	information technology
ITIC	Information Technology Industry Council
LLC	Inductor-inductor-capacitor
PEM	Proton exchange membrane [electrolyzer]
PFC	Power factor correction
POI	Point of interconnection
PSPD	Positive-sequence phasor domain
PSU	Power supply unit
PUE	Power usage effectiveness
SCR	Short-circuit ratio
SOEC	Solid oxide electrolytic cell [electrolyzer]
SST	Solid state transformer
THD	Total harmonic distortion
TPU	Tensor processing unit
UPS	Uninterruptible power supply
VFD	Variable frequency drive

Introduction

A majority of the surge of electrical load growth for the next five years is from large data centers and other large loads such as manufacturing and hydrogen electrolysis facilities, all sized from 500 MW up to 1.5 GW per site. To understand the risks that large loads impose on power system planning and operation, power system operators and transmision owners must understand how these large loads behave and how they can affect the power system's operational security. This will require both detailed models of the large loads and a thorough understanding and representation of load-to-grid interactions and behaviors. This report discusses the behaviors, capabilities, and limitations of these types of large loads; demonstrates the need to adopt interconnection requirements for large load facilities; and covers large loads' capabilities (and limitations on their capabilities) to reduce grid reliability risks.

Many of the types of large loads connecting to the power system today are power electronics–based and have faster control mechanisms and new dynamics relative to more traditional motor-driven loads. Four specific types of large load problems affect grid operational security:

- **The speed of large load operation.** Today's large power electronic loads can have the ability to ramp within seconds, and the introduction of a power electronics interface between the load and the grid presents new modeling complexities. Given the size and the increasing number of large loads, power system planners and operators need to understand and address how these new complexities can impact the grid. For example, the high variability of artificial intelligence (AI) workloads and other data center behaviors could exhaust balancing reserves.
- **Electronic interface and trip sensitivity**. Many power electronic loads are highly responsive to faults and other grid conditions and have protection settings that can cause multiple loads to reduce demand or trip simultaneously, leading to over-frequency and/or voltage problems that have the potential to cause cascading outages.
- **Difficulty in forecasting load behaviors in operational time frames.** Many large loads operate in unpredictable ways, making their electricity usage hard to predict and react to at the grid level. Cryptocurrency miners are price-responsive, which affects short-term and longer-term load forecasts. AI training facilities can start and stop training at a moment's notice, which can be very difficult to predict.
- **Potential for harmful harmonics and oscillations.** The control methods of power electronics can cause oscillations to originate from both data centers and cryptocurrency mining. These facilities also exacerbate the risk of increased harmonic emissions due to their power electronics interfaces.

Table 1 summarizes the characteristics of load designs, behaviors, and capabilities for four types of traditional and new large loads, listing the key features that matter to power system engineers. (This summary is a general reference and will not hold true for every possible load design.) This report focuses principally on data center, cryptocurrency mining, and hydrogen loads, while Appendix A includes information on industrial and transportation loads.

TABLE 1

Summary Characteristics of Large Load Power Characteristics

	Traditional Data Center	Artificial Intelligence (AI) Data Center	Cryptocurrency Mining	Hydrogen
Facility function	Provides compute and storage for enterprise and cloud applications	Provides capability for AI training and inference	Performs cryptocurrency mining	Produces hydrogen for use in industrial processes and energy applications
Main compute/process technology	Central processing units (CPUs), graphics processing units (GPUs), high-speed memory, hard drives	GPUs, specialized chips for training, such as tensor processing units (TPUs)	Application-specific integrated circuits (ASICs)	Proton exchange membrane (PEM) electrolyzers, alkaline electrolyzers, solid oxide electrolyzers (SOEC)
Back-up power	Yes	Yes	No	Yes
Uninterruptible power supply	Yes	Not needed if using a checkpoint save process, but may be installed in mixed traditional/AI data centers	No	No
Load profile	Stable: requires high uptime to meet contractual requirements	Highly variable for AI training applications, moderately variable for AI inference	Stable, unless certain pricing conditions cause demand reduction	Stable, but can be driven by contractual obligations, electricity or hydrogen pricing, or availability of renewable energy supply
Flexible electricity use	Typically not flexible	Potentially flexible	Flexible	Potentially flexible
Ride-through capability	Designed to traditional computer voltage limits	Designed to traditional computer voltage limits	Designed to traditional computer voltage limits	No unified standard (jurisdiction dependent)

Source: Energy Systems Integration Group.

The Energy Systems Integration Group's Large Loads Task Force produced a set of four interrelated reports on large load interconnection performance requirements, of which this report is a part.[1] This report also informs the ESIG Large Loads Task Force report *Large Load Modeling for Dynamic Studies: Current Practices and Recommendations*.[2]

- The main report, *Large Load Performance Requirements: Current Practices and Recommendations*, provides guidance for system operators and utilities as they develop interconnection performance requirements for large loads tailored to their own jurisdictions.
- The report *Large Load Disturbance Events* highlights a set of recent disturbances in which large loads played a noticeable role, either as contributors to the initiating conditions or as factors exacerbating an initial disturbance and complicating a recovery.
- The report *Reliability Impacts of Large Loads* provides a comprehensive technical analysis of the reliability impacts of large loads related to dynamic stability, identifying critical vulnerabilities and proposing mitigation strategies for system planners, grid operators, and load developers to address these dynamic stability risks.

This report helps power system operators and transmission owners understand how new large loads behave and how they can affect the power system's operational security, and thus understand the risks that large loads impose on power system planning and operation and gain insight into their mitigation.

[1] See also *Large Load Performance Requirements: Current Practices and Recommendations*, *Reliability Impacts of Large, Power Electronics–Interfaced Loads*, and *Large Load Disturbance Events*. All four reports can be found at https://www.esig.energy/working-groups/large-loads.
[2] https://www.esig.energy/reports-briefs/large-load-modeling

Data Centers

Data centers provide computing and storage capabilities to end users, which could be the facility owner or other clients that lease space in the facility. Large data center types include:

- **Enterprise data centers,** which serve a single end-use customer.
- **Multi-tenant/co-location data centers**, facilities where multiple customers can lease space and install their own equipment or lease equipment already on site.[3] The data center owner is responsible for maintenance and facility management. Since these facilities serve users beyond the facility's owner, they typically have uptime and latency requirements.
- **Hyperscale data centers** typically serve a single customer but are much larger than traditional data centers (generally 500 MW or larger. These typically provide infrastructure for cloud and AI applications.
- **AI data centers** are focused on training large AI models, especially large language models. They use the same internal equipment as other data centers, with different configurations and a much higher percentage of graphics processing units (GPUs) or other advanced accelerator technologies for AI training.

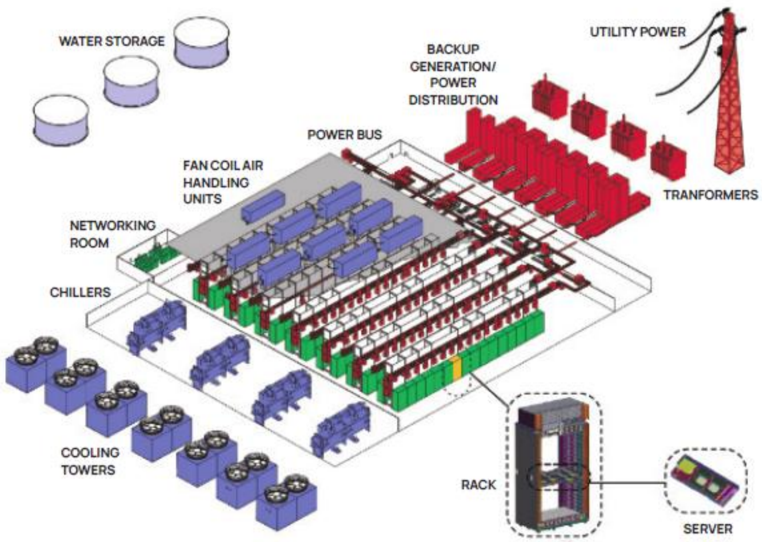
Facility Design

Data centers vary in design based on their specific functions, but they generally share common core subsystems and layouts. The overall facility is typically a warehouse containing server racks, cooling, network, power distribution, back-up power, and water storage (Figure 1). Large data centers are built in modular blocks that can be expanded depending on the amount of computing power needed.

[3] This meaning of co-location of data centers, where users can rent space to install their servers without running their own data center, is distinct from co-location as a pairing of a large load with associated generation facilities.

FIGURE 1

Core Components and Generic Design of a Large Data Center



Source: L. A. Barroso, U. Hölzle, and P. Ranganathan, *The Datacenter as a Computer: Designing Warehouse-Scale Machines,* 3rd ed. (Springer Cham, 2022), https://link.springer.com/book/10.1007/978-3-031-01761-2.

While these subsystems are consistent across most facilities, the design specifics and energy consumption of the subsystems and components may vary as a function of each facility's purpose, vintage, and ownership (IEA, 2025). The most common subsystems in data centers are:

- **Servers and compute equipment**: Devices like central processing units (CPUs), GPUs, and tensor processing units (TPUs) that processes data
- **Data storage:** Devices responsible for storing data such as hard drives
- **Networking**: Equipment facilitating connectivity between servers within the data center and the outside world
- **Cooling and environmental controls**: Equipment responsible for managing temperature and humidity; this infrastructure depends on facility design end use, and climate
- **Uninterruptible power supply (UPS), batteries, and back-up power generators**: Equipment that supplies power inside the data center site during outages
- **Other infrastructure**: Services such as lighting, office equipment, and others

Understanding the facility design is critical to do accurate data center modeling and to understand their expected behavior and grid impacts. When modeling data centers for load projections, the power usage effectiveness (PUE) can help evaluate overall project energy consumption based on the available information about its compute equipment.

The PUE describes a data center's overall energy efficiency and is a ratio that describes how much energy is used for computing equipment versus all other equipment, such as cooling and lighting. A facility's PUE depends on the technology used and its purpose. Numerous estimates of PUE exist, with the Uptime Insititute reporting that average data center PUE is around 1.56 and has not improved since 2020 (Donnellan et al., 2024). However, Google reported an average PUE of 1.09 for its sites in its latest data from 2025 (Google, 2025).

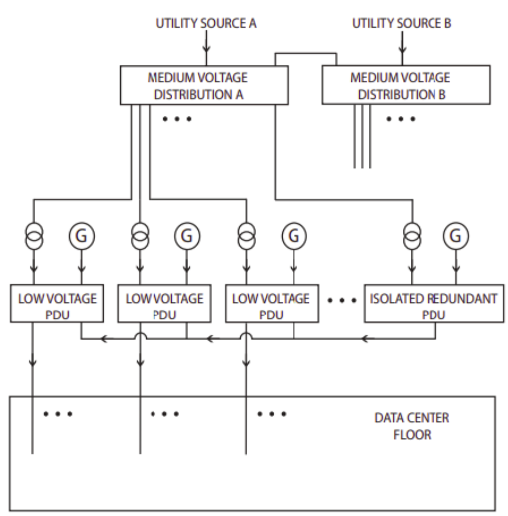
$$PUE = \frac{Total\ Facility\ Energy}{IT\ Equipment\ Energy} = 1 + \frac{Non - IT\ Facility\ Energy}{IT\ Equipment\ Energy}$$

Internal Power Distribution System

Large data center loads connect to the transmission system, and the data center's internal power distribution system is the primary interface between the data center and the grid. This power distribution and conversion equipment and design must be understood to be accurately modeled and used for the grid impact studies that simulate and predict the data center's response to grid disturbances. Large load modeling is discussed in detail in the ESIG Large Loads Task Force report, *Large Load Modeling for Dynamic Studies: Current Practices and Recommendations*.[4]

Figure 2 shows a conventional radial power distribution system within a Google data center. In the power distribution system the incoming power feed from the utility goes through a step-down transformer to medium voltages around 11 kV to 15 kV (not pictured). That medium-voltage network distributes power throughout the building. Back-up power generators are paralleled to the medium-voltage buses, and a transformer steps the medium voltage down to low voltages around 400 V AC for distribution to server racks.

FIGURE 2
Conventional Radial Power Distribution System at Google Data Center



Source: L. A. Barroso, U. Hölzle, and P. Ranganathan, *The Datacenter as a Computer: Designing Warehouse-Scale Machines,* 3rd ed. (Springer Cham, 2022), https://link.springer.com/book/10.1007/978-3-031-01761-2. https://link.springer.com/book/10.1007/978-3-031-01761-2

[4] https://www.esig.energy/reports-briefs/large-load-modeling

Most of the compute and power equipment in a data center is designed to the Information Technology Industry Council (ITIC) curve (Figure 3), which specifies the range of voltages that equipment can withstand without damage or shutting down. In the event that the power system voltage from the grid to the data center falls outside the ITIC curve, the data center will shift to back-up power, either from UPS units or a battery back-up unit, before shifting to back-up generation. If the data center has no back-up power solutions then the power supply unit will typically follow the ITIC curve and disconnect from the grid. However, as noted in the ESIG Large Loads Task Force report *Large Load Performance Requirements: Current Practices and Recommendations*:

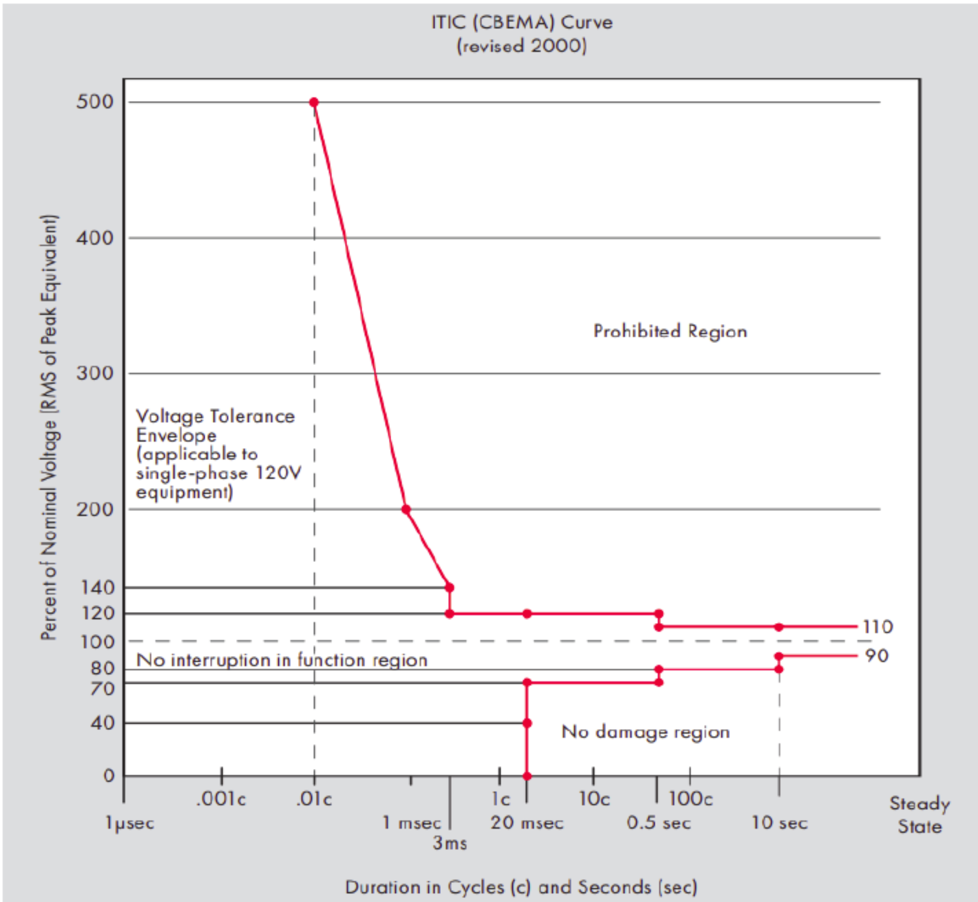
> The ITIC curve is not intended to maximize the ability of large load facilities to remain grid-connected during grid disturbances. As a consequence of this focus on internal reliability needs, a large load's voltage ride-through capability may not be sufficient to ensure grid reliability, as voltage tolerances are misaligned with expected power system disturbance profiles and timing. Dedicated large load voltage ride-through curves, focusing on the grid's reliability needs, are necessary to ensure that connecting facilities are able to ride through faults on the power system.[5]

In addition, since data centers and vendors often keep the exact specifications of information technology (IT) equipment confidential, it is hard for grid operators to predict the voltage level at which the data center will disconnect from the grid.

[5] https://www.esig.energy/reports-briefs/large-load-interconnection-performance-requirements.

FIGURE 3

ITIC Voltage Curve Specifying the Range of Voltages That Equipment Can Withstand Without Damage or Shutting Down



Source: Information Technology Industry Council.

Importance for the power system: When performing studies that model the impact of a large load loss, the ITIC curve is a good proxy to use for the data center's response to faults. However, the modeler should request better information from the interconnecting data center on its actual fault-ride-through capabilities.

The following sections discuss UPS units, which provide short-term back-up power to IT infrastructure in the data center and are a major component of the power distribution system, and data center internal power distribution system architectures. Both affect power system interactions and modeling.

Uninterruptible Power Supplies

UPS units are a critical part of a data center, and modeling UPS behavior is critical for accurate grid impact simulations, both in positive-sequence phasor-domain (PSPD) and electromagnetic transient (EMT) simulations. The runtime of the UPS also influences how long the IT equipment can support the data center after system disturbances (such as nearby grid faults) before it transitions the data center onto its back-up power supply. While on the UPS, the data center can return to grid-connected operation relatively quickly (in seconds), but once the facility transitions to back-up generation, it will take much longer to reconnect to the grid.

Most enterprise, multi-tenant, and hyperscale data centers use UPS units to provide short-term redundancy in the event of a power failure. These data centers require high energy supply availability and use a UPS between the server and the grid. Facilities primarily focused on high-performance computing, such as national laboratories or AI training centers, typically do not rely on UPS units for IT equipment because they use a checkpoint process to save their work (NERC, 2025a). With the checkpoint process for AI training loads, if there is a power failure, then some training progress is lost but the training run can be restored to the last saved snapshot. Any training since the snapshot and the power failure is lost. For high-performance computing, checkpoints are saved in the simulation process so that if an outage occurs, the last known good state (which is saved in UPS unit–backed storage) can be restored and the process can continue. In these designs, the cost of the UPS units does not justify the benefits in saved training progress or simulations (Salim, 2026).

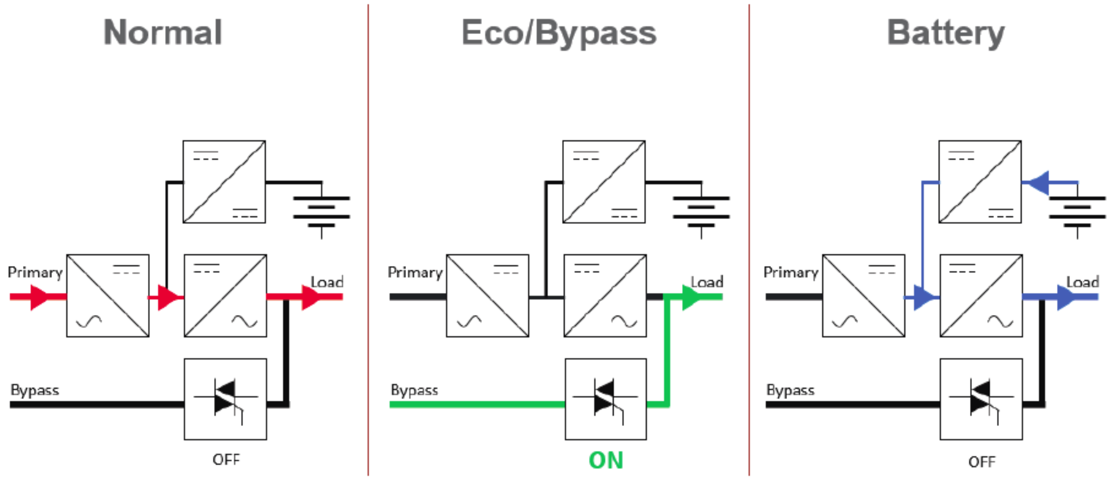
The runtime of a UPS will depend on the its sizing relative to the host facility's total load. Some common UPS runtimes are (Hurley, 2020):

- Hyperscale facilities: 1 to 2 minutes
- Cloud and co-location facilities: 5 minutes
- Facilities serving financial customers: 10 to 15 minutes

UPS Design

The most common UPS design is a double conversion UPS. These UPS units have three modes: normal, eco/bypass, and battery, shown in Figure 4. In normal operation, the power flows through the primary conversion equipment to the load. In eco/bypass mode, the power flows through a static bypass; this mode is used to reduce losses in normal operations. In battery mode, power is supplied to the load from the battery. All of the equipment in the UPS is used to ensure seamless transfers.

FIGURE 4
Double Conversion UPS Operating Modes



Source: B. Ross, "UPS Fault Dynamics," Slide deck from presentation at ERCOT Large Load Working Group meeting, July 11, 2025, https://www.ercot.com/files/docs/2025/07/10/PNNL_ERCOT_UPS_Presentation_Outline.pptx.

Inside a double conversion UPS, the front end of the UPS—the primary side in Figure 4—controls the voltage on the DC bus between the primary and load side and maintains the power factor on the grid side. The back end—the load side of the UPS—controls the output to the data center bus, buffering it from grid side distortions. The battery is always connected to the DC bus ready to supply power and charging using the DC-DC converter.

If the grid voltage drops (as a result of a nearby fault, for example), the DC bus voltage drops as well. The IT load is a constant power load so that as the voltage drops, the current increases. If the rectifier current limit is reached, it begins to block and shut down, causing the data center to disconnect from the grid and switch to UPS-supplied power. Then, as the voltage recovers on the grid, the UPS may switch the load back to the grid and begin recharging (Mitra, 2025).

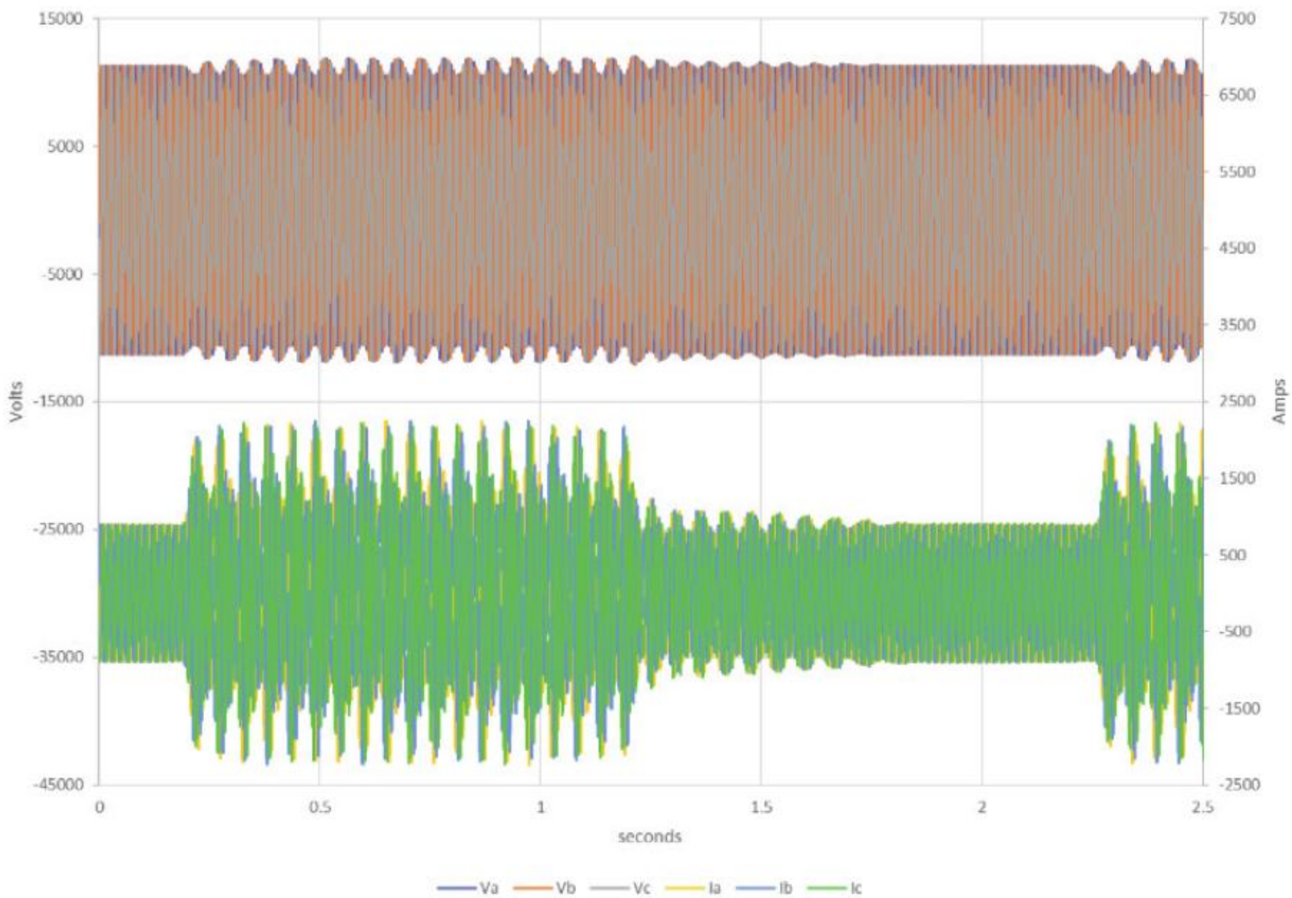
UPS Units' Contributions to Oscillations

Inside data centers, computing processes can cause repeated load fluctuations that can cause forced oscillations on the grid beyond the facility meter, and the power electronics interface can contribute to resonance and converter-driven instability. Both types of oscillations can interact with nearby generators and cause uncontrolled generator tripping. UPS units have the potential to contribute to these oscillations. In the ESIG Large Loads Task Force survey of large load entities, several reported that oscillations have occurred at their data centers.

One load entity performed testing of its servers and identified load-dependent rectifier stability issues. At the facility, they turned up the full set of servers to full load, then idled them in a synchronous manner. When the load was high, the UPS units (40 units arranged in 20 pairs) would begin oscillating (Figure 5). The exact point where oscillations began depended on the feed inductance and load on the units. They were able to stop the oscillations by removing one of the pairs of UPS units from the utility feed, thereby decreasing the effective impedance seen by the other pairs. They were able to replicate this behavior both in models and in testing with the UPS manufacturer by including representative inductance in the UPS utility connection. The oscillations were corrected by reducing the loop gain of the rectifier power factor correction (PFC) circuit to acquire the necessary phase margin. While these oscillations were only seen on the data center facility side of the point of interconnection (POI), as more large data centers come online, these oscillations could have more frequent and harmful impacts on the power system.

FIGURE 5

Uninterruptible Power Supply Rectifier Oscillations



UPS rectifier oscillations as observed on a data center medium-voltage bus attributed to the loop gain of the rectifier power factor correction circuit.

Source: Energy Systems Integration Group.

In these tests the typical onset of the oscillations occurred as load was increased in the data center. In one case, however, the oscillations were caused by the utility's removal of one of two transformers that had been operating in parallel, increasing the effective inductance feeding the UPS pairs.

This load entity also reported another oscillation that was due to an unstable power supply PFC stage. These manifested at both a site with a UPS and a site without one and at much higher frequency than the UPS issue, in the 9^{th} to 11^{th} harmonic range. The load entity was able to work with the vendor to modify the rectifier control to eliminate the issue.

Importance for the power system: Given that UPS units inside a data center are capable of causing oscillations on the power system, thorough modeling of the facility is needed to screen for potential oscillatory behavior, and models must reflect the UPS PFC stage where oscillations originated. These events highlight the need to enhance real-time situational awareness by monitoring data centers with high-resolution monitoring devices like phasor measurement units (PMUs) or digital fault recorders, to complement PMUs placed at the POI between the large load and the bulk power system.

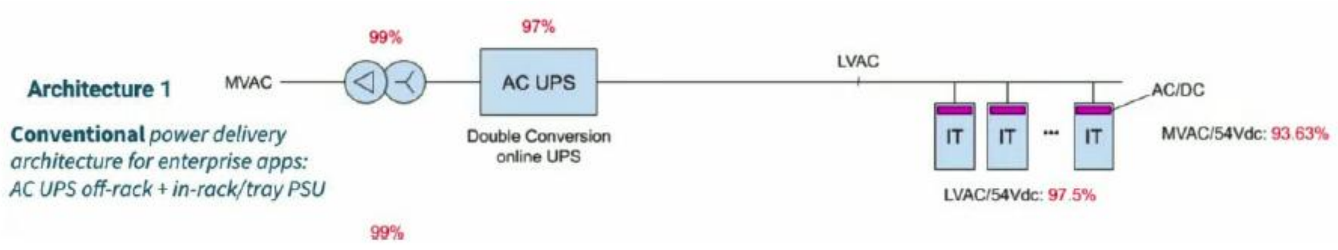
Power Distribution Architectures

The power distribution and conversion equipment within the data center is one of the main pieces of equipment that drives the facility's dynamic response to grid disturbances. The design or architecture of a data center's internal power distribution systems and its associated control systems can have numerous impacts on the external grid, including the ability of a data center to ride through voltage disturbances (the failure to ride-through could negatively impact grid stability) and the potential for a data center to become the source of oscillations on the power system from mistuned control systems. The architecture can also impact how the data center is modeled for positive-sequence phasor domain (PSPD) and EMT simulations. Inside the data center these architectures are duplicated, fanning out from the grid through the internal power distribution system to the parallel IT equipment.

The Open Compute Project has specified several current and future server power infrastructure designs, documented below in architectures 1 through 4 (Figures 6, 7, 8, and 12) (Li, Ravikumar, and Peabody, 2024a). Each figure shows an architecture and the efficiency of each form of power conversion in red. In the current conventional power delivery design in service today (Figure 6), a three-phase transformer with primary side medium voltage and secondary side 416/480/600 V AC steps the voltage down. The power then flows through a double conversion UPS and is converted to 54 V DC with an in-rack power supply unit (PSU). However, data center designs continue to evolve to support higher power density racks and higher distribution voltages, discussed below.

FIGURE 6

Conventional Power Distribution Architecture



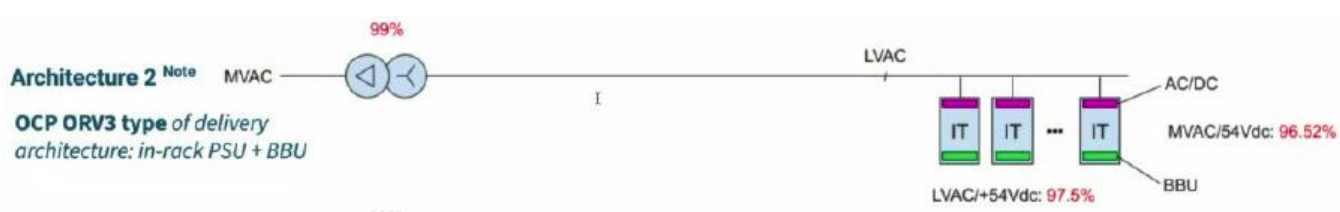
The conventional power distribution architecture inside data centers features power delivered through a transformer to a double conversion online UPS and then the server racks, where the low-voltage AC at 416/480 V is converted to 54 V DC.

Source: X. Li, K. G. Ravikumar, and C. Peabody, "+/- 400 Vdc Rack for AI/ML Applications," Slide deck from presentation at Open Compute Project Global Summit, October 15-17, 2024, https://drive.google.com/file/d/1BCiJ5XJg1bbhF7I_BKvnSzjoUClUdGk2/view.

The Open Compute Project ORV3 architecture (Figure 7) consists of a three-phase transformer with primary side medium voltage and secondary side 416/480/600 V AC stepping the voltage down, which is then distributed to the server racks and converted to 54 V DC. The battery back-up unit is combined with the in-rack PSU and battery back-up unit.

FIGURE 7

Open Compute Project ORV3 Architecture

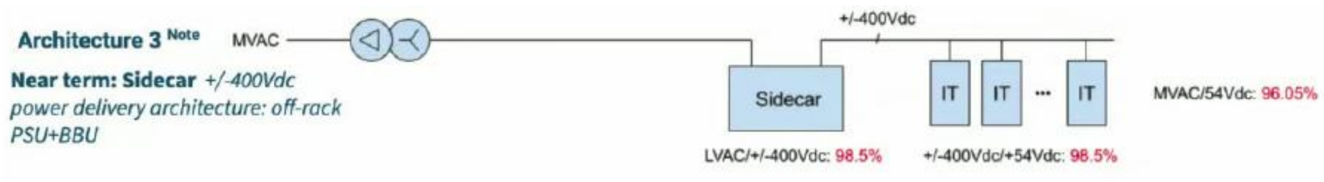


The Open Compute Project ORV3 power distribution architecture for data centers consists of power delivered through a transformer to the server racks as low-voltage AC at 416/480 V, where it is converted to 54 V DC. Each server rack also has a battery back-up unit to provide back-up power.

Source: X. Li, K. G. Ravikumar, and C. Peabody, "+/- 400 Vdc Rack for AI/ML Applications," Slide deck from presentation at Open Compute Project Global Summit, October 15-17, 2024, https://drive.google.com/file/d/1BCiJ5XJg1bbhF7I_BKvnSzjoUClUdGk2/view.

The growth in power demands from machine learning workflows means that future server racks could use more than 500 kW per rack by 2030. To support this, 400 V DC is needed and can support up to 1 MW per rack. To support the need for higher rack power densities, new power distribution architectures are adopting 400 V DC and 800 V DC for internal distribution. Two distinct 400 V DC architectures are shown in Figures 8 and 9. In Figure 8, incoming power is stepped down to 416/480/600 V AC, then converted to 400 V DC via a "sidecar" system, which distributes the 400 V DC to servers, where a DC-DC converter reduces the voltage to 54 V DC.

FIGURE 8
Sidecar Architecture



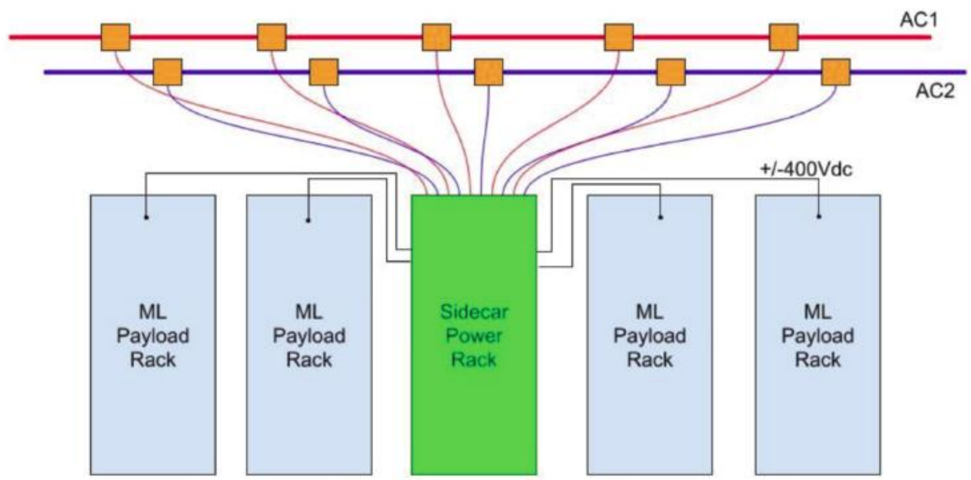
To meet the growth in power demand from machine learning workflows, the power distribution architecture features power delivered through a transformer to a sidecar. The sidecar, which features the power supply and battery back-up unit, takes in low-voltage AC and converts it to 400 V DC. Then, 400 V DC is distributed to the servers, where it is converted to 54 V DC.

Source: X. Li, K. G. Ravikumar, and C. Peabody, "+/- 400 Vdc Rack for AI/ML Applications," Slide deck from presentation at Open Compute Project Global Summit, October 15-17, 2024, https://drive.google.com/file/d/1BCiJ5XJg1bbhF7l_BKvnSzjoUClUdGk2/view.

In Figure 9, the 400 V DC architecture supplies 400 V DC with the sidecar system where the power infrastructure is split from the compute infrastructure in the rack (Iyengar and Huffman, 2025). This allows all of the power infrastructure to be centralized in a single server rack, freeing up space in other racks for more compute infrastructure. This also reduces the amount of power conversion that must be performed, increasing efficiency and making cooling easier. To ease the transition to the sidecar systems, the legacy 48 V DC racks are supported by the new 400 V DC power sidecar; the sidecar power rack can be deployed with or without battery back-up units.

FIGURE 9

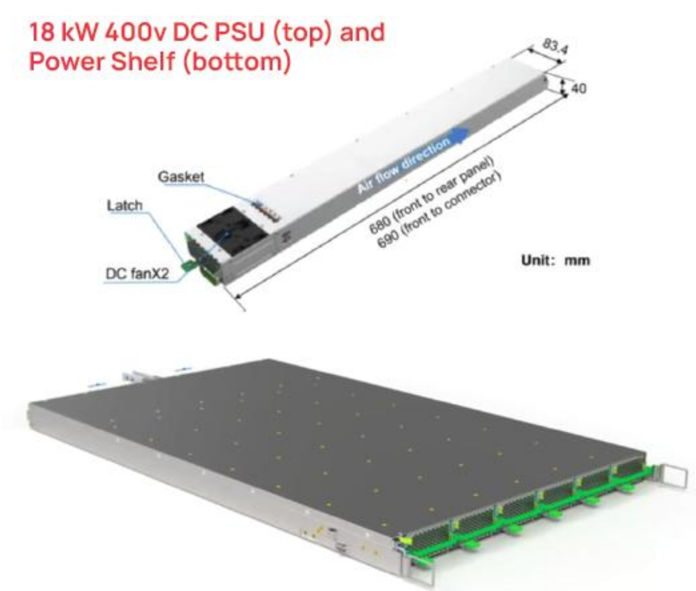
Sidecar with Power Infrastructure and Compute Racks



Source: X. Li, K. G. Ravikumar, and C. Peabody, "+/- 400 Vdc Rack for AI/ML Applications," Slide deck from presentation at Open Compute Project Global Summit, October 15-17, 2024, https://drive.google.com/file/d/1BCiJ5XJg1bbhF7l_BKvnSzjoUClUdGk2/view. https://drive.google.com/file/d/1BCiJ5XJg1bbhF7l_BKvnSzjoUClUdGk2/view?usp=drive_link?usp=drive_link

A new power supply is needed to use emerging 400 V DC in the data center power system. A 400 V DC PSU can handle 400 V DC from the sidecar and then convert the voltage to 50 V DC for the IT rack. Each PSU can supply up to 18 kW. This ultimately enables up to 400 kW per rack. Six PSUs are then combined into a single power shelf, shown in Figure 10 (Sapre and Sras, 2025).

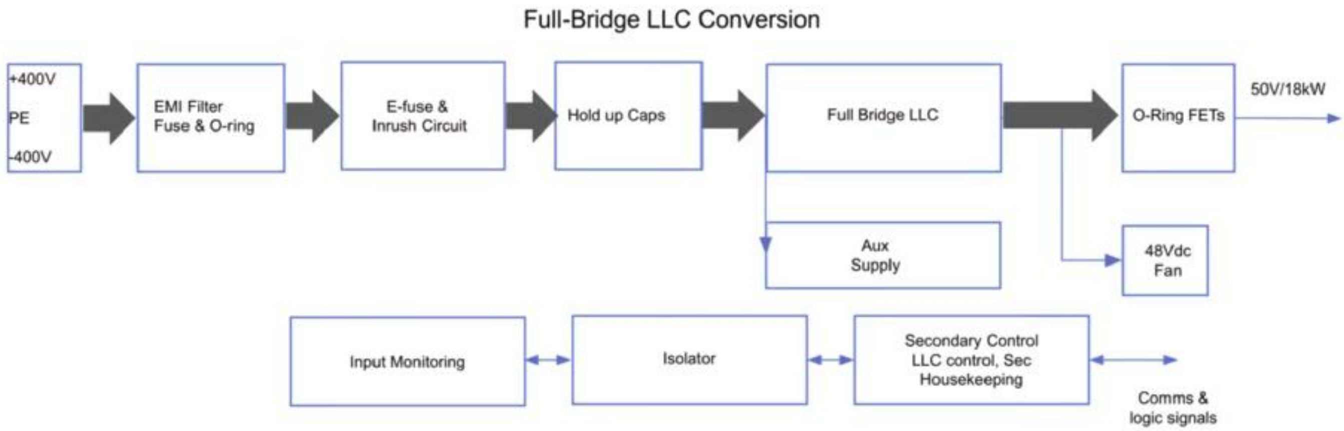
FIGURE 10



Source: S. Sapre and Z. Sras, "OCP Presentation 18kW DC PSU 100kW DC Power Shelf," presentation, July 9, 2025.

Each power supply uses a full-bridge LLC (inductor-inductor-capacitor) conversion. A full-bridge converter is a type of resonant DC-DC converter widely used in high-efficiency power supplies. It leverages a full-bridge topology and an LLC resonant circuit to achieve efficient power conversion with soft switching techniques. In the block diagram for the PSU (Figure 11) (Sapre and Sras, 2025), the full-bridge converter transforms the input DC voltage (400 V) to a lower-output DC voltage (50 V) with high efficiency.

FIGURE 11
400 V DC PSU Block Diagram

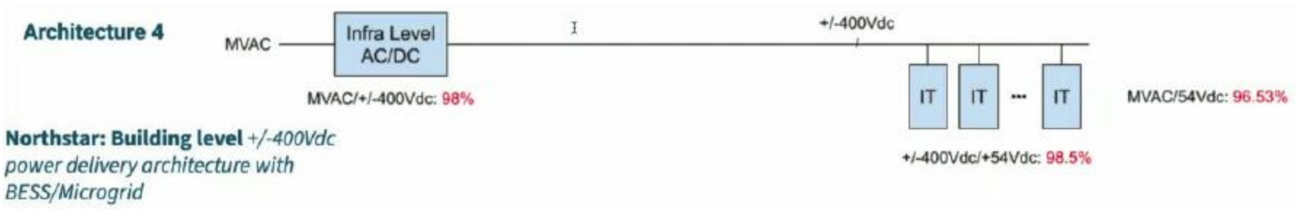


Source: S. Sapre and Z. Sras, "OCP Presentation 18kW DC PSU 100kW DC Power Shelf," presentation, July 9, 2025.

Importance for the power system: The 400 V DC power supply is a key component of the sidecar system and needs to be understood for proper system modeling.

The second 400 V DC (Northstar) architecture uses a solid state transformer (SST) to convert medium-voltage (11 kV, 13.8 kV, or 35 kV) AC to 400 V DC. Shown in Figure 12, this design could also utilize a rectifier or other conversion method in place of the SST. With facility-level 400 V DC power distribution, less conversion is needed compared to the sidecar architecture, further increasing efficiency. The 400 V DC is distributed to the servers where a DC-DC converter steps the voltage down to 54 V DC.

FIGURE 12
Building-Level Architecture



To fully support growing server power demand, the power distribution architecture of the future will feature building-level AC/DC power conversion using a rectifier or solid state transformer that converts the power delivered to 400 V DC. This power will be delivered to the servers where it is converted to 54 V DC.

Source: X. Li, K. G. Ravikumar, and C. Peabody, "+/- 400 Vdc Rack for AI/ML Applications," Slide deck from presentation at Open Compute Project Global Summit, October 15-17, 2024, https://drive.google.com/file/d/1BCiJ5XJg1bbhF7l_BKvnSzjoUClUdGk2/view.

In addition to the proposed 400 V DC architectures, Nvidia has proposed an 800 V DC architecture for future machine learning workflows with high power demands. Figure 1 in Nvidia's technical blog in October 2025 shows a conventional internal power distribution network as used today in the top diagram and Nvidia's proposed 800 V DC architecture in the lower diagram (Srivastava and Petty, 2025).[6] This new proposed architecture should support power densities up to 1 MW. Like the 400 V DC architecture, it utilizes centralized AC/DC conversion infrastructure and a sidecar design, using a higher voltage to minimize energy losses. High AC voltage (13.8 kV) from the grid is converted directly to 800 V DC at the data center perimeter using rectifiers or a solid state transformer. This approach reduces the PSUs needed (although they are still present for data center reliability). The 800 V DC is transmitted via busways through data halls to equipment rows. Overcurrent protection is integrated at boundaries between the power room to hall, hall to row, and row to IT racks. This set-up carries 85% more power through the same conductor size, thus lowering resistive losses. The IT rack receives 800 V DC via two-conductor feeds, and inside the rack, DC/DC converters step it down to 54 V/12 V. Eliminating AC/DC conversion elements allows for higher-density configurations and improved cooling efficiencies compared to traditional AC/DC conversion stages.

The power distribution and conversion equipment in the data center is one of the main pieces of equipment that drives the facility's dynamic response to grid disturbances, because the equipment needed to meet emerging ride-through requirements is primarily installed on its power distribution system. As data centers evolve, new models must be developed to model the sidecar power conversion system, and the infrastructure-level AC/DC converter (which may be an SST). Additionally, the dynamic response may also be influenced by the 400 V DC (or 800 V DC) to 54 V DC DC-DC converter, which may need to be modeled depending on the exact system architecture.

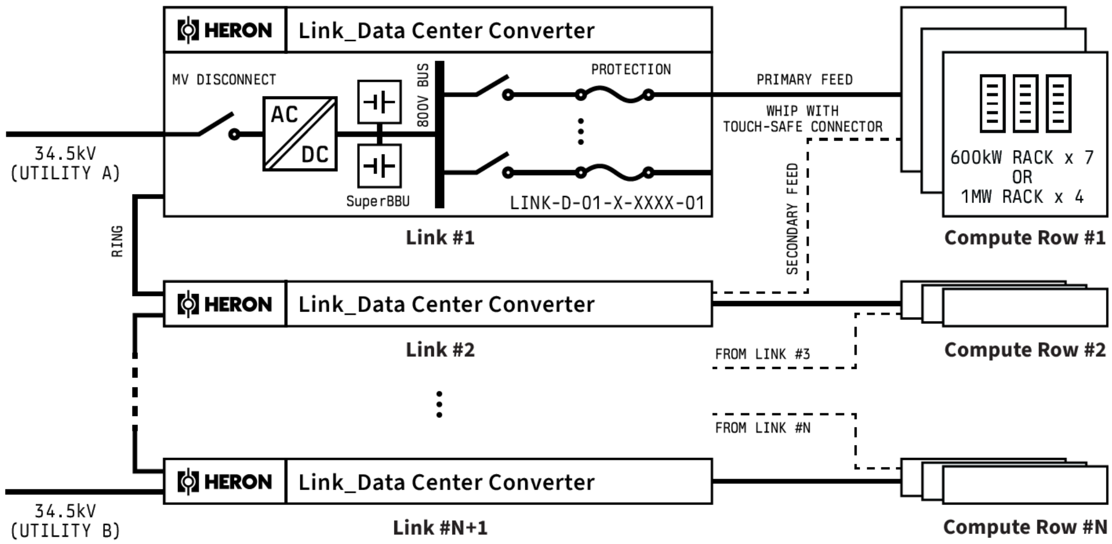
In emerging designs, SSTs are often envisioned as the site-level AC/DC power conversion method to supply the 400/800 V DC network. As shown in Figure 13, SSTs use a power electronic interface to transform AC voltages,

[6] https://developer.nvidia.com/blog/building-the-800-vdc-ecosystem-for-efficient-scalable-ai-factories/

convert AC to DC, and provide active regulation of voltage and current. When coupled with energy storage, SSTs can be used to meet emerging requirements such as ride-through requirements. SSTs will eliminate multiple levels of power conversion and reduce energy losses.

FIGURE 13

Conceptual Diagram of a Heron Link Solid State Transformer for Data Center Applications



Source: Heron Power, "Heron Link: Unlocking Efficient Power Conversion for Gigascale AI," white paper, October 2025, https://heronpower.com/resources/unlocking-efficient-power-conversion-for-gigascale-ai.

Importance for the power system: The design of power distribution systems has numerous impacts on the power system.

- Power conversion equipment design determines the capabilities and limitations of the equipment to withstand fault-induced low voltages. Increasing inverter overload capability may help mitigate fault-induced voltage problems and improve fault ride-through.
- To properly simulate the data center, modelers will need to create PSPD and EMT models for the new architectures including the sidecar, SST, and other power conversion equipment.
- Power conversion equipment could become a source of troublesome harmonics that may need study and mitigation.

Back-Up Generation

Data centers require continuous reliable power to maintain operations, since even brief outages can lead to significant financial losses. Back-up generation ensures uninterrupted services by providing redundant power sources that activiate during grid disturbances (e.g., voltage fluctuations). Once grid voltage falls to or below a certain threshold, a data center will switch over to its UPS (see the section "Uninterruptible Power Supplies"). Depending on its size, the UPS's support may last for a few seconds to tens of seconds. If grid voltage has not recovered during this time, most data centers will use on-site back-up generation (or battery energy storage systems) for longer-term power support. Data centers will typically transfer to back-up generation if triggered by their protection schemes following a disturbance. Back-up generators have a minimum run-time, usually around 20 to 30 minutes; the data center cannot immediately reconnect once it has transferred to back-up power.

Importance for the power system: The minimum run-time requirements on back-up generators will limit facilities' ability to return immediately following a grid disturbance if the data center transfers to back-up generation.

Compute Systems

At the heart of a data center are its compute systems. The nature of the compute systems inside a data center will influence its electric load profile. Data center compute systems use CPUs, GPUs, TPUs, and other compute chips and use data storage, memory, and connectivity to complement those compute systems. All together, these systems process all the user requests to perform computations, store and retrieve data, and manage customer workloads. Each type of compute chip has a different power profile and intensitity. These are further influenced by the nature of the facility's work that the compute process supports (e.g., AI training, AI inference, cloud computing).

Historically, CPU clusters performed simulations and processing. These systems were known as high-performance computers. In the 1990s, GPUs became more prevalent as consumer 3D graphics became more common. GPUs perform highly parallel calculations, whereas CPUs are more general-purpose and carry out calculations in sequential order. Thus, GPUs excel at solving massively parallel linear alegbra equations, which are the foundation of machine learning models. Most AI training facilities are composed primarily of GPU clusters consisting of rack-mounted GPUs (see image of Nvidia Blackwell Ultra GPU racks at Nvidia (2025)).[7]

Importance for the power system: Data centers with CPU-based compute generally have a more stable load profile due to both their workloads and sequential processing design. In comparison, today's more common GPU-based compute has a more variable load profile due to its workloads and parallel processing capabilites.

Cooling Systems

Because semiconductors generate so much heat, data center cooling and environmental systems are critical to manage temperature and humidity inside the facility. Cooling functions can consume as much as 30% to 40% of a data center's total electricity use.[8]

[7] https://nvidianews.nvidia.com/news/blackwell-ultra-dgx-superpod-supercomputer-ai-factories

[8] The percentage of data center power usage by cooling systems also depends on the data center archtitecture and construction date. Newer data centers are more efficient than older ones as the facility PUE improves. Thirty to forty percent is a top-of- range number for cooling systems as a share of total data center load, because data center owners are motivated to reduce non-compute energy consumption to improve profit.

The main types of data center cooling systems are:

- Air cooling
- Liquid cooling
- Immersion cooling
- Evaporative cooling
- Free cooling
- Hybrid cooling

For cooling systems that have fans or pumps, nearly all are driven by variable frequency drives (VFDs) or electronically commutated (EC) motors. VFDs can be programmed to have different ride-through capabilites than other equipment in the data center, influencing the entire facility's ride-through capabilites. This must be recognized in facility power modeling. The ride-through capability could be better or worse, depending on the facility's programming. Both types of motors are power electronics–based, but EC motors may have a greater harmonic output than VFDs (Dynamic Air, 2026).

Cooling systems are critical to the reliable operation of the data center. If a cooling system trips offline (e.g., due to ride-through issues), the data center compute systems will need to shut down to protect all of its equipment and operations.

Importance for the power system: Data centers with cooling systems driven by VFDs may be susceptible to ride-through issues with both their VFDs and IT equipment. Cooling system failure would trigger a shutdown of the data center if redundant cooling systems fail. Data center cooling system load affects the data center PUE and its overall energy consumption, so understanding the nature of the cooling system will enable more accurate forecasting of both short-term responses to grid events and longer-term load profiles.

Operational Characteristics

Conventional Data Center Load Profile

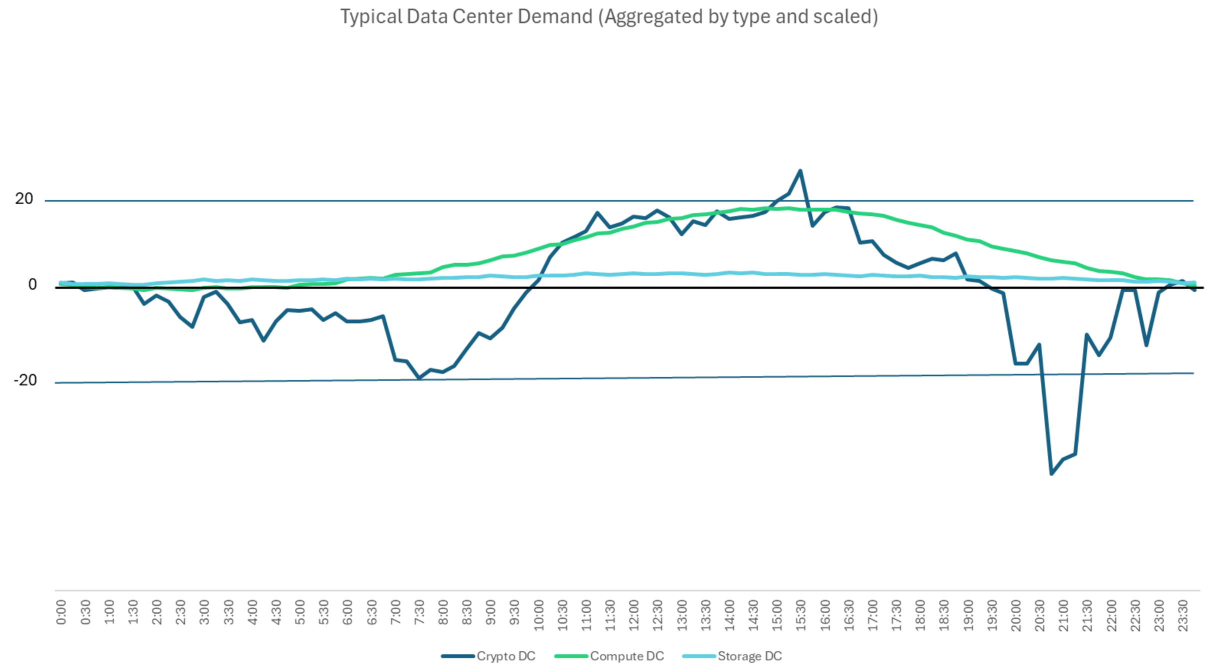
A data center's load profile is driven by its primary function. Enterprise, cloud, and multi-tenant data center load profiles are likely to show more stable analytical operations and energy use. Data center owners/operators rarely disclose their load profiles due to the confidential nature of the information.

The Electric Reliability Council of Texas (ERCOT) analyzed all loads classified as cloud storage and cloud compute data centers of all sizes and found distinct patterns in the load profile (Figure 14). A compute data center's load profile displays the behavior of a conforming load, which follows the overall system's demand pattern, while storage data centers have a more consistent, less variable load profile.

Importance for the power system: When considering how data centers influence load forecasts, the end-use purpose of the data center (e.g., compute, storage, cloud, or AI training or inference) must be considered.

FIGURE 14

Data Center Demand Patterns in ERCOT, Which Vary by the Data Center's Purpose



Source: S. Morris, "Data Center Inside ERCOT: Observations and Patterns," Slide deck from presentation at ERCOT Large Load Working Group meeting, August 14, 2025, https://www.ercot.com/files/docs/2025/08/11/LLWG_DataCenterObs.pptx.

AI Workload Load Profiles

AI data centers, purpose-built facilities designed to handle the unique demands of AI workloads, exhibit fundamentally different electrical and operational characteristics compared to conventional data centers or other types of electrical loads. The primary distinction lies in their reliance on GPU clusters, which are optimized for matrix multiplications—a critical operation for neural network training and inference. These GPU clusters operate in large, synchronized groups to ensure the efficient execution of computational tasks. This synchronization yields unique power consumption patterns for differing AI model workloadsAI training data center load profiles are much more variable than conventional data centers, whereas AI inference data center load variability is somewhere between AI training data centers and conventional data centers. Table 2 compares AI workloads with traditional computational loads.

TABLE 2
Comparison Between AI Workload and Traditional Computational Load

Aspect	Traditional Computing Workload	AI Workloads
Power density	10–15 kW per rack	30 kW per rack or more
Load variability	Stable, predictable	Highly variable, bursty
Computational intensity	Moderate	Extremely high
Load profile transition points to new consumption patterns	Few; mostly stable	Training/fine-tuning/inference

Source: Energy Systems Integration Group.

AI workload power consumption has three components: compute power for intensive calculations, which can reach up to 90% of peak during parallel tensor operations; memory power for accessing model weights and activations, which is significant due to billions of internal model variables that store what the model has learned (known as parameters); and transfer power for data movement, particularly in distributed training (Li and Li, 2025). These components combine to produce rapid load transients, where surges in memory and compute coincide during attention steps or inference requests, amplifying peak demands and distinguishing AI from slower CPU-based loads. Unlike conventional data centers, where power usage is relatively firm, rapid load transitions in AI data centers are driven by the training, fine-tuning, and inference workloads. Each type of AI work shows the following distinct load profiles:

- **Training**: Training involves processing large-scale models, often requiring the entire data center's GPU clusters to operate in a synchronized manner. During training cycles, power consumption may follow periodic patterns dictated by the software algorithms and hardware configurations at the rack level. These cycles can last for hours or days, depending on the complexity of the model. However, computational failures or job completion can cause sudden drops in power consumption (sometimes spread over multiple data centers, if all are used for the same big compute job), even down to idle (inactive) operation.

- **Fine-tuning**: Fine-tuning, an intermediate activity between initial model training and inference, involves refining a pre-trained model for a specific application or dataset. Although less resource-intensive than full-scale training, fine-tuning still demands substantial GPU usage, often in bursts rather than continuous high loads. This yields a more dynamic power profile with fluctuating energy demands and thermal output relative to traditional enterprise or multi-tenant data centers that typically run more non-AI-related misssion-critical workloads.

- **Inference**: Inference tasks typically involve smaller models and shorter computational cycles. While the power fluctuations are less pronounced compared to training, they still exhibit peaks and drops as GPUs alternate between computationally intense and less intense phases.

Although at the single GPU level the power might jump within microseconds, the power consumption of the supporting PSU will be limited for up to ~50 milliseconds. This is due to the PSU's input current remaining at the lower level of power consumption for 1 to 2 cycles before it increases due to the PSU capacitor and control loop timing providing/absorbing energy. This means that the overall ramp will be limited to up to ~50 milliseconds even though the GPU energy consumption can ramp more quickly (Li and Li, 2025).

Most AI training uses a backpropagation algorithm (see below) operates in the following cycle, which dictates the overall load profile of a GPU (Bergmann and Stryker, n.d.):

1. Forward pass: Here, the algorithm makes predictions based on the training data and power demand spikes during this period as the GPU processes the training data.
2. Loss function calculation: Next, the algorithm measures the errors from the earlier predictions, and power demand decreases.
3. Reverse/backwards pass (also known as backpropagation): The algorithm calculates the partial derivative of the loss function, and power demand increases.
4. Gradient descent: Lastly, the algorithm adjusts the model weights, and power demand decreases.

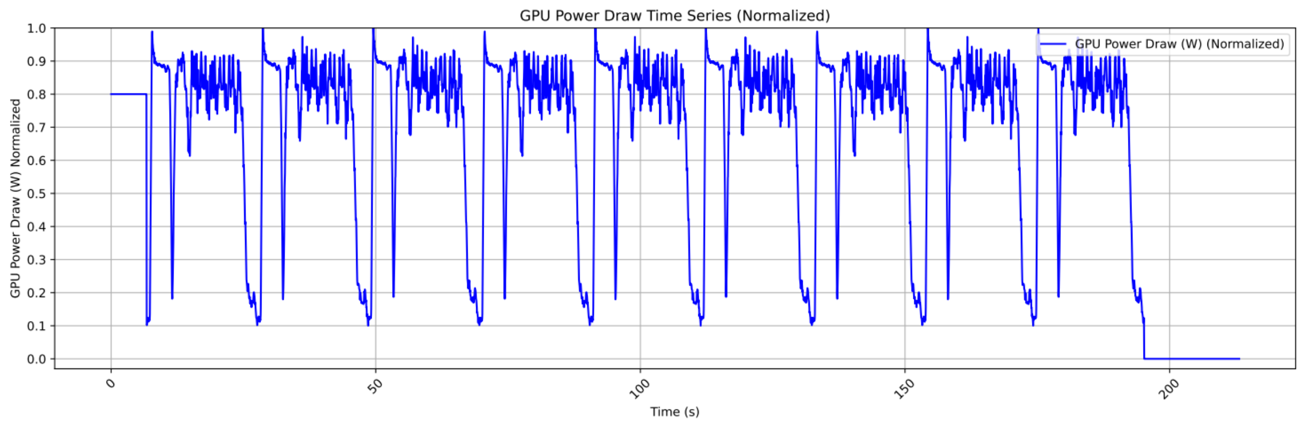
Since AI training processes require all of the machines to be operational, if a single machine fails, the entire training operation will stop. This could drop up to 30% to 50% of the active load within microseconds. When this happens, the training job will auto-reconfigure and restart. This stop and restart sequence can happen tens of times per day.

AI training data centers operate at a baseline of around 60% to 70% of peak electricity demand to maintain readiness. The training requirements demonstrate operational variations, with steady high-energy consumption followed by intensified memory activity and bursts of full load followed by partial idling. This creates sawtooth profiles with stochastic fluctuations from micro-batching or background analytical tasks.

The AI training load profile is characterized by repetitive behavior as the program executes. Figure 15 provides an example of an AI training load profile, showing significant variability in the GPU load profile.

FIGURE 15

AI Training Server GPU Load Profile



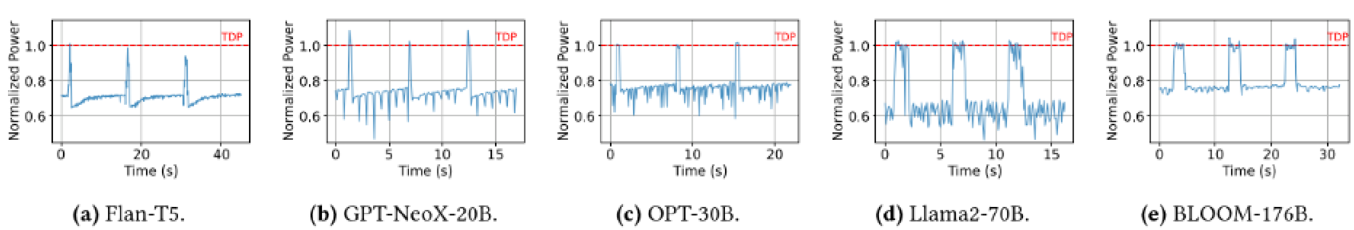
Normalized GPU power draw during a test AI training run demonstrating the signifcant variability in the GPU load profile.

Source: E. Choukse, et al., "Power Stabilization for AI Training Datacenters," preprint, arXiv, August 21, 2025, https://doi.org/10.48550/arXiv.2508.14318.

AI data centers also perform inference, which involves processing multiple requests and exhibits bursty operational behavior. Power consumption for AI inference features low baseline idle power use, punctuated by each work request's arrival and demand pattern, yielding oscillations between idle and near-peak as requests arrive in bursts. Individual inference operations may use less power than training, but frequent transitions and latency needs for the aggregate inference workload may cause rapid spikes and drops of as much as 80% to 90% power reductions in under a second. Figure 16 shows normalized power use patterns (i.e., adjusted to show the relative magnitude of each pattern rather than total use for each pattern examined) for GPU-based inference across multi-second periods. This shows how the inference power profile differs from the power profile during GPU-based training in Figure 16 above.

FIGURE 16

Examples of Different AI Models' Workload and Power Use Patterns during GPU Inference



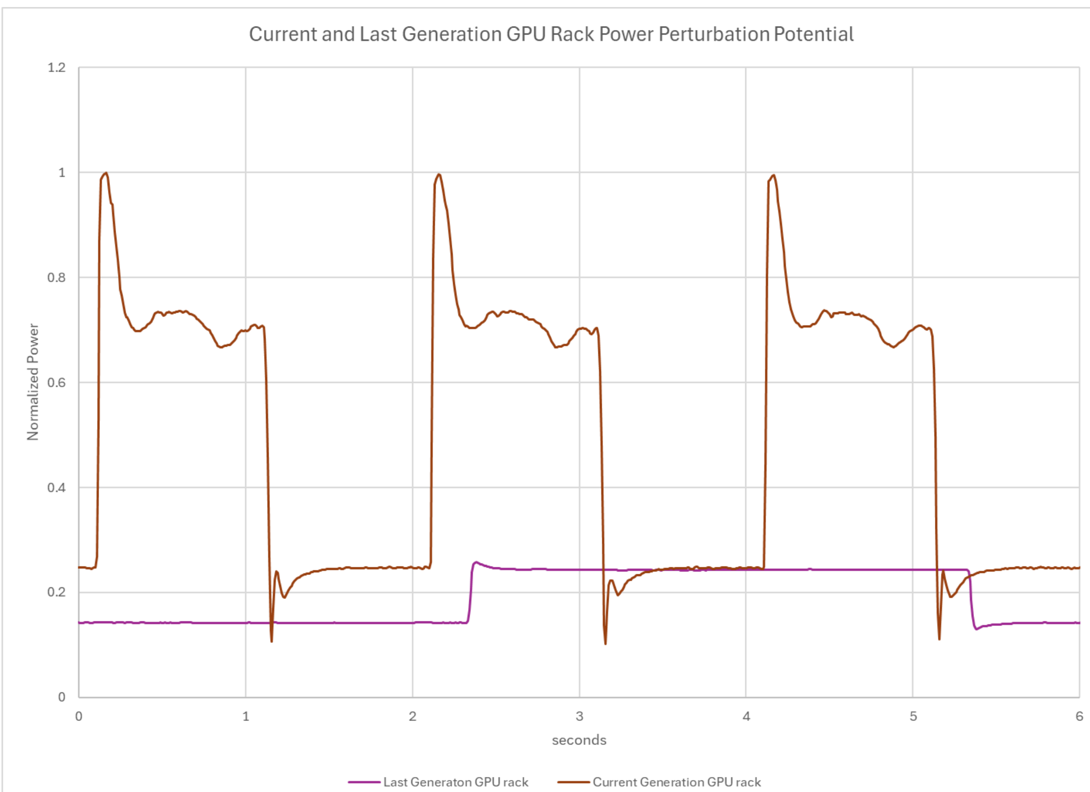
Source: P. Patel, et al., "Characterizing Power Management Opportunities for LLMs in the Cloud," in *Proceedings of the 29th ACM International Conference on Architectural Support for Programming Languages and Operating Systems, Volume 3,* 2024, https://doi.org/10.1145/3620666.3651329.

With highly variable and repetitive AI training workloads, AI data centers can experience repeated load fluctuations which, because of the large size of the sudden change in grid demand, can cause forced oscillations on the external grid that can interact with and damage nearby generators. This potential harm depends on the frequency spectrum and magnitude of the oscillations (NERC, 2025b).

As the percentage of GPU-driven data center load increases, the amount and variability of load will increase relative to the amount of more stable and predictable load from other grid users. This variability is mostly driven by the use of liquid-cooled GPUs. Figure 17 shows GPU rack-level normalized power consumption data for prior-generation versus current GPUS, showing how power use variability significantly increases from the last to the current generation of GPUs; the architecture of the newer GPUs causes higher power-use variation that poses new risks to the grid. The data in the figure are a test profile; neither the period nor the duty cycle represent a real load, but the relative variability is accurate.

FIGURE 17

Comparison of Rack-Level GPU Power Consumption



GPU rack-level power perturbation potential can vary significantly between generations of GPUs, showing that advacements in GPU architecture can have new impacts on the power system.

Source: Energy Systems Integration Group.

Given the rapid innovation in GPU architectures and AI algorithms, future AI data center load profiles may look very different than what is seen today. This makes it difficult for grid operators and transmission owners to assess the potential impact of the AI data center load fluctuations on the grid and nearby equipment and design appropriate mitigations to these impacts. Additionally, GPU or rack level workload profiles when aggregated across multiple server racks and data center modules can produce a different load pattern at the data center point of connection to the grid.

Importance for the power system: As AI models increase in size and complexity, their energy demands grow disproportionately. Scaling from millions to billions, or even trillions, of parameters significantly amplifies the computational intensity of both training and inference. For instance, training a model like GPT-3, which has 175 billion parameters, is estimated to require hundreds of megawatt-hours of electricity, depending on hardware efficiency (Patterson et al., 2021). The relationship between AI model size and power use is not linear. Increasing AI model size often leads to non-linear increases in energy use due to longer training durations, greater memory bandwidth demands, and larger datasets (OpenAI, 2018).

Moreover, the widespread deployment of inference contributes substantially to an AI model's overall energy footprint, particularly for widely deployed AI models with high query volumes (Schwartz et al., 2020). While training on one AI model is a one-time event, new AI models are always being trained. Inference activity is a continuous process—especially in real-time applications—which results in persistent high power consumption over the AI model's lifecycle. Inference energy consumption can match or even exceed training costs when models are integrated into production environments (Strubell, Ganesh, and McCallum, 2019).

These scaling trends toward ever larger model sizes place increasing strain on the underlying grid infrastructure and have raised sustainability concerns, prompting efforts to enhance AI model efficiency, the ongoing development of energy-optimized hardware, and carbon-aware workload scheduling (Thompson et al., 2022; Google, 2023).

The variability of AI training loads will continue to increase as the number of GPUs in the data center increases compared to overall data center load. This is exacerbated by routine failures of equipment in the data center, causing the training run to stop and restart multiple times throughout the day. Understanding backpropagation can help system modelers develop better AI facility load profiles that operators can use to integrate the data center load more effectively and manage the grid more reliably. Extreme, high-speed load variability must be limited or managed during real-time system operations. Increased load variability also poses challenges for managing balancing reserves and grid frequency. Furthermore, if frequency of power fluctuations due to AI training causes oscillations in frequencies that coincide with resonant modes on the grid, that could excite inter-area oscillations or cause torsional oscillation issues at nearby synchronous generators.

Distributed AI Model Training

In recent years, deep learning (multi-layered neural network) models have significantly increased in size, complexity, and data requirements. Modern architectures such as GPT-style models, vision models, and diffusion models now contain hundreds of billions of parameters and require petabytes of training data. Training these large-scale models on a single GPU, or even a single server, has become infeasible due to limitations in memory capacity, compute throughput, and acceptable training duration.

Distributed deep learning (DDL) has emerged as a critical strategy to overcome these challenges. It enables the distribution of training workloads across multiple GPUs, servers, and even geographically dispersed data centers. This approach reduces training time and allows organizations to scale their operations more efficiently by spreading AI tasks across multiple data centers. DDL ultimately divides the training workload across multiple compute nodes; these could be GPUs within a single server, across multiple servers, or across data centers globally. However, DDL introduces new challenges related to communication, synchronization, and energy consumption patterns, which could have significant implications for reliable power system operation.

Importance for the power system: If a data center engaged in DDL experiences a fault and trips offline, it will send a signal to other data centers engaged in the same DDL job to pause training. Thus, if a model being trained across multiple data centers suddenly stops at one data center due to a grid fault, communication failure, or coordinated disconnection, then multiple data centers may reduce operations and reduce electricity demand as their training jobs also pause. This coordinated load reduction across multiple multi-GW data centers in the same balancing authority or interconnection could pose a serious risk to the stability of system frequency if it is not planned for and mitigated effectively. This abrupt imbalance between generation and demand could trigger over-frequency protection schemes, force generating units to reduce output or trip, and, in extreme cases, lead to cascading generation tripping and blackouts. As data centers become more geographically concentrated and energy-intensive, their collective behavior—especially during synchronized or unplanned shutdowns—must be anticipated and accounted for in system stability planning and grid operation protocols. Estimating the potential size and electrical scope of such events is a moving target and thus remains a challenge.

Load Flexibility and Fluctuations

Historically, small data centers (100 MW and below) did not impose a large burden on the grid because they were too small to move frequency noticeably relative to other loads. These small facilities were not designed or operated to provide demand flexibility to the grid. Data centers are not price-sensitive, as the return on compute is much higher than the return on load flexibility. However, as more very large load facilities connect to the grid, more grid operators and regulators are requiring that large loads develop and use more flexibility tools such as the ability to reduce their load, shift computational loads to a different region, or accept curtailments in the event of extreme system conditions.

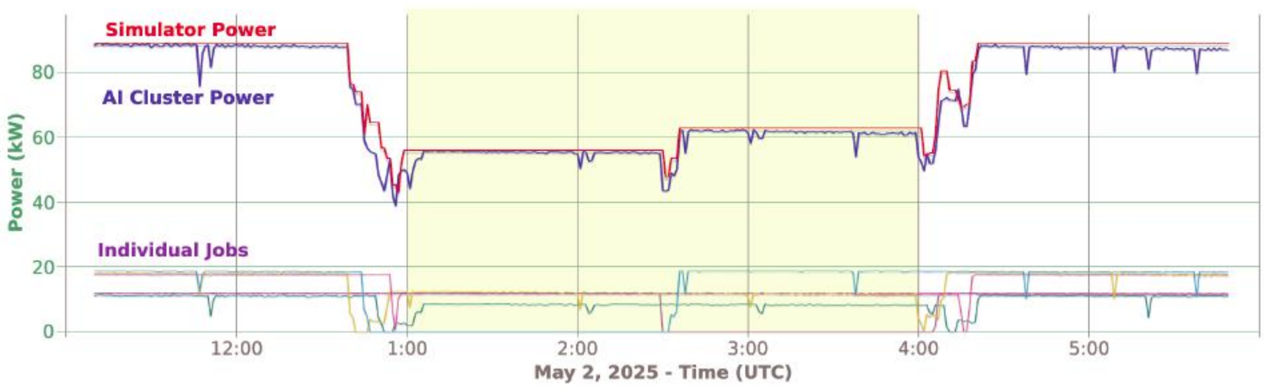
AI training workloads are a promising way to gain energy use flexibility from some data centers. Before AI inference models can be used commercially, they must be trained in an offline process. Since AI training happens prior to model release, it does not face the same exacting performance requirements from enterprise customers—it can be run, then stopped and restarted without much loss or complaints from internal customers. These make model training a good target for operational flexibility, but because inference operations are highly variable, they challenge grid stability and themselves drive system flexibility needs.

Not all data centers can easily develop operational flexibility. Certain facilities like multi-tenant/co-location data centers have contractual uptime requirements from their customers. In such facilities, the customer, rather than the data center operator, runs the center’s compute workload and therefore its electricity demand; therefore, the customer will be crucial to potential demand flexibility development.

There are several approaches to providing flexibility and power smoothing and mitigate the rapid power fluctuation inherent in modern data centers, particularly those driven by AI training workloads. These include both software- and hardware-based approaches.

There have been demonstrations of load flexibility using software to manage GPU clusters for AI training. Emerald AI, Nvidia, the Salt River Project, and EPRI completed such a small pilot demonstration using job pausing and dynamic frequency scaling in which the GPUs' clock speed and voltage were adjusted to manage performance and power consumption, shown in Figure 18.

FIGURE 18
Demonstration of Load Flexibility



This diagram shows software-based load flexibility implemented by EmeraldAI, Nvidia, the Salt River Project, and EPRI using job pausing and dynamic frequency scaling in which GPUs' clock speed and voltage were adjusted to manage performance and power consumption.

Source: P. Colangelo, et al., "Turning AI Data Centers into Grid-Interactive Assets: Results from a Field Demonstration in Phoenix, Arizona," 2025, https://doi.org/10.48550/arXiv.2507.00909.

Other software-based approaches to load shaping and flexibility include Google's compiler-based approaches, balancing training jobs across time for optimized power flow management (Gan and Ranganathan, 2025). This form of compiler-based mitigation can help smooth load fluctuations. Non-critical jobs can be routed to other data centers on different power systems to managing larger-scale fluctuations and provide demand response (Mehra and Hasegawa, 2023).

Hardware-based solutions include using capacitor-based systems inside the data center to provide a fast-response, high-efficiency solution. These devices can smooth transients that could otherwise propagate to the grid and cause voltage instability or trigger protective relays. They store energy using capacitors and inject or absorb power to smooth the fluctuations in the data center load profile. Unlike traditional batteries, which degrade under frequent high-rate cycling, capacitors and supercapacitors provide millions of charge/discharge cycles and ultra-fast response times, making them ideal for bridging the gap between local PSUs and facility-level UPS units. They can integrate into rack- or facility-level architectures. Several vendors offer capacitor-based products to manage the variability of data center operations and corresponding electric loads, including:

- Siemens Energy's E-STATCOM (Siemens Energy, 2026)
- Eaton's XLHV, a rack-mounted supercapacitor bank (Eaton, 2025)
- Delta Electrics' Power Capacitance Shelf, a rack-mounted bank of lithium-ion capacitors(Delta Electronics, 2025)

Batteries can also provide hardware-based flexibility and power smoothing at different points inside the data center. Batteries offer a fast response and provide greater energy capacity to support longer periods of power smoothing and steeper load fluctuations. The Tesla Megapack is an example of a battery unit that can provide internal power smoothing for data centers around the clock (Tesla, 2025).

Importance for the power system: Certain data centers such as co-location facilities, where multiple tenants occupy a single facility, can be contractually required to maintain minimum uptime for their customers, and these facilities are less likely to be flexible in response to grid operator requests. AI training data centers experience large power fluctuations that must be managed, but they offer an opportunity for innovative load flexibility and demand response applications. Grid managers will need to develop load flexibility options in partnership with large load entities. While capacitor-based solutions and energy storage installed at the data center may mitigate power fluctuations, other less capital-intensive methods such as code optimization and job management may also address this issue. Capacitor-based solutions could be used locally as rack-mounted solutions for AI training load variability.

Cryptocurrency Mining

Cryptocurrency (or crypto) is decentralized virtual currency. The most well-known cryptocurrency is Bitcoin. Bitcoin stores its transactions on a decentralized blockchain where a peer-to-peer network maintains a record of all transactions. The process of creating cryptocurrencies, can use either a “proof of work” method or “proof of stake” method. Proof of work mining is much more energy-intensive, as computers solve mathematical problems to unlock new coins. This is a competitive effort, the first person to solve the problem gets the Bitcoin so miners are incentivized to deploy ever more powerful computers that can solve the problems faster than others. With a proof of stake method, new coins are distributed to those “staking” their coins to validate transactions and to enter the “lottery” for obtaining more cryptocurrency. The more coins that are staked, the higher the probability that one part gets the new coinsSince proof of stake methods are based on a “lottery” type system versus a compeititon on how who solve mathermatical problems faster there is no need to build ever more compute thus rendering proof of stake more energy efficient than proof of work methods. The process of generating cryptocurrencies like Bitcoin is known as mining. As more cryptocurrencies are mined, more computational resources are needed to produce each new crypto unit. Cryptocurrency miners are a growing load class on the grid that can affect power system dynamics and overall energy consumption. In ERCOT, for example, there have been numerous events where many crypto miners have failed to ride through grid disturbances and dropped offline in unison, causing a frequency perturbation across the ERCOT system. Crypto miners have a high degree of load variability because they are highly price-responsive and willing to participate in compensated demand response programs, so it is difficult to accurately forecast their short-term electricity demand.

Facility Design

Cryptocurrency mines generally use one of two design architectures, either a warehouse-style facility or self-contained containers that can be dropped off at a site. Both designs feature the same core components: the mining computer rigs, power distribution systems, cooling, and other auxiliary equipment.

Cryptocurrency Miners

A typical cryptocurrency mining facility operates thousands of individual mining units. Miners are typically run application-specific integrated circuits (ASICs) for computing. ASIC miners are single-phase devices plus a neutral connection.

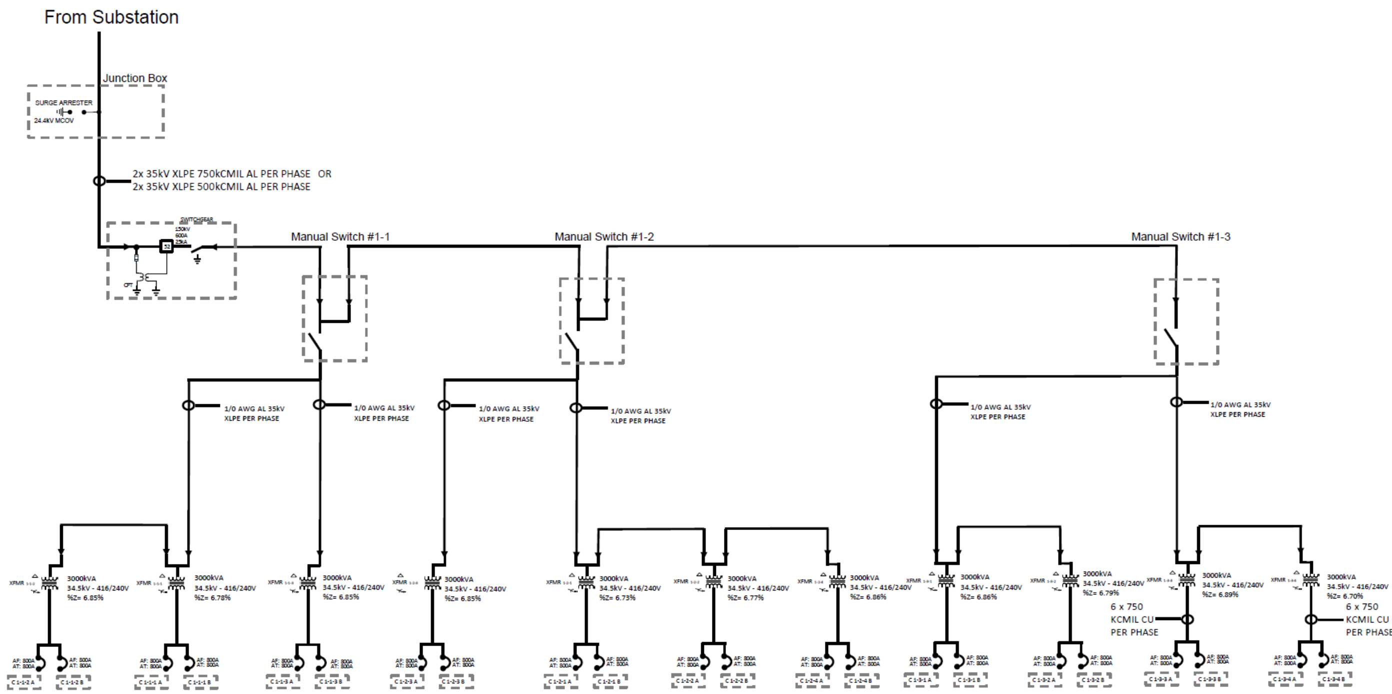
Common mining rigs include Antminers such as the S19 or S21, Microbt M66, and Whatsminer.

Power Distribution System

A large cryptocurrency facility connects to the transmission system through a transformer stepping the transmission line voltage down to a distribution voltage such as 34.5 kV. As shown in Figure 19, transformers step the voltage down even further to 416/240 V inside the facility for use by the miners.

FIGURE 19

Sample One-Line Diagram for Cryptocurrency Mining Facility



Source: MARA.

Since cryptocurrency mining is primarily an arbitrage play where miners turn energy into money, cryptomining facility operators seek to minimize their cost. Therefore, most cryptocurrency miners do not use UPS units or back-up generators because that equipment would increase total costs.

Importance for the power system: Cyptocurrency miner models must represent the component mining rigs accurately to capture the facility's operations and electricity use effectively.

Cooling Systems

Miners use several different cooling methods including air cooling, hydro-cooling such as single-stage immersion cooling, or dielectric cooling. A hydro-cooled miner is self-contained in a skid but it can shut down if the temperature gets too high. The cooling fans for both air- and hydro-cooled miners can be driven by VFDs, which have limited ride-through capability.

Several crypto miners have reported their facilities tripping offline for various reasons. In one case, the facility tripped three times, and the root cause was found to be the low-voltage ride-through capabilities of the VFD (where voltage in one phase dropped by 60% from the grid). This was resolved by changing the VFD parameters and was only applicable for single-phase immersion miners.

Importance for the power system: In addition to IT equipment failing to ride through a grid disturbance, cooling systems driven by VFDs are also susceptible to ride-through failures. If the cooling system fails to ride through or shuts down for another reason, the entire cryptocurrency mining facility will shut down, exacerbating the impact of the intial disturbance. VFDs in cooling systems at cryptocurrency mining facilities should be reviewed and set to ensure they do not trip off during normal faults.

Auxiliary Equipment

In a cryptocurrency mining facility around 1% to 5% of electrical load is classified as auxiliary load that is needed to ensure the safe operation of the facility and requires continuous power delivery. This load is essential for health and safety and can include cooling and security-related loads. A cryptocurrency mining facility may also have capacitors for power factor correction and relays for internal monitoring and monitoring at its POI to the grid.

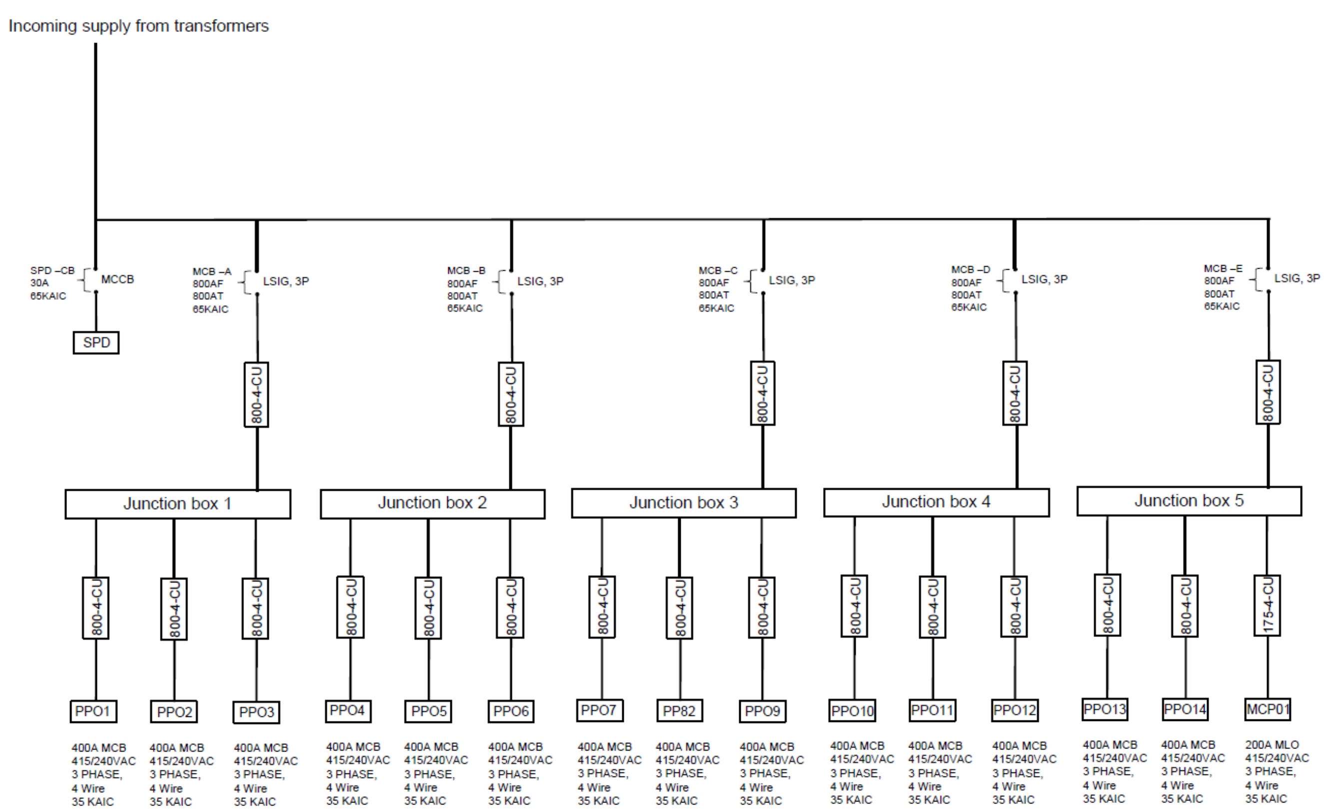
Sample One-Line Diagrams

The sample one-line diagrams in Figure 20 show details on the layout of a cryptocurrency mining facility, including the mining rigs and cooling system layouts.

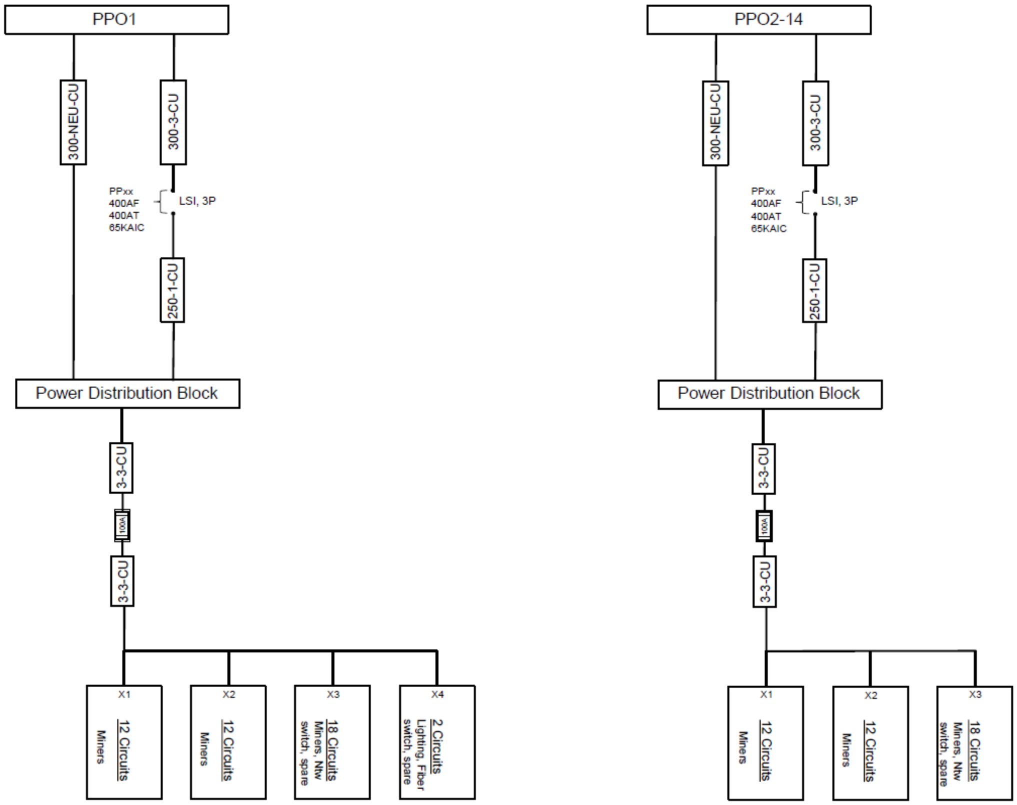
FIGURE 20

Sample One-Line Diagrams of a Cryptocurrency Mining Facility Showing the Hardware Connections

A



B


PPO1
300-NEU-CU
300-3-CU
PPxx
400AF
400AT
65KAIC
LSI, 3P
250-1-CU
Power Distribution Block
3-3-CU
100A
3-3-CU
X1
12 Circuits
Miners
X2
12 Circuits
Miners
X3
18 Circuits
Miners, Ntw
switch, spare
X4
2 Circuits
Lighting, Fiber
switch, spare
PPO2-14
300-NEU-CU
300-3-CU
PPxx
400AF
400AT
65KAIC
LSI, 3P
250-1-CU
Power Distribution Block
3-3-CU
100A
3-3-CU
X1
12 Circuits
Miners
X2
12 Circuits
Miners
X3
18 Circuits
Miners, Ntw
switch, spare

C

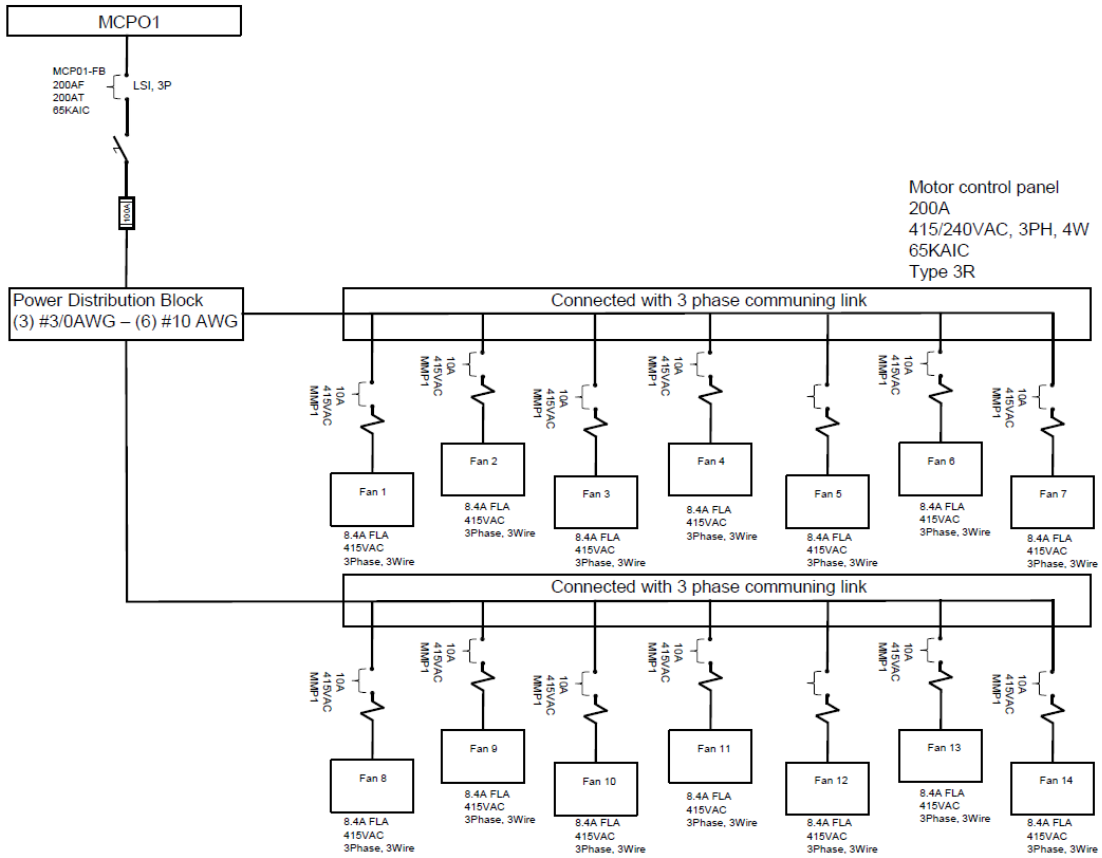


Figure A depicts a sample cryptocurrency mining facility electrical system. Figure B depicts a sample cryptocurrency mining facility electrical diagram for wiring between the main junction boxes to the end-use loads. Figure C depicts a sample cryptocurrency mining facility electrical diagram for wiring from the junction box to a motor control panel to cooling fans.

Source: MARA.

Operational Characteristics

Crypto miners try to minimize equipment and energy costs and adjust their operations in response to real-time electricity prices. Typically, crypto mining is marginal with around a 20% to 40% investment rate of return. The miner's goal is to get as high a hash rate as possible, as this will enable the miner to obtain more cryptocurrency.[9] Cypto mining faces high commercial risk due to the volatility of cryptocurrency values and

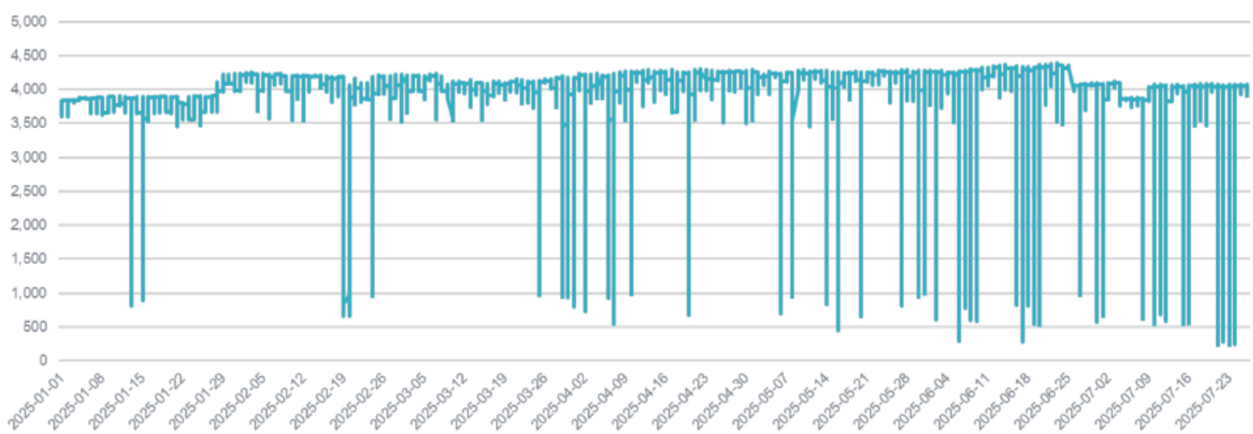
[9] Hash rate is a measure of cryptocurrency mining computational power and can be interpreted as the speed of crypto mining. The higher the hash rate, the faster (more) cryptocurrency is mined. The hash rate is a function of the mining hardware's ability to solve the mathematical puzzle that underlies Bitcoin mining.

energy prices, so miners aim to operate with as little non-mining equipment as possible to earn the highest possible rate of return as possible on the overall facility investment.

Load Profile

Under normal grid conditions, crypto miners show little variability in demand because they want to operate as much as possible to maximize crypto production and income (Figure 21). Miners aim to have the lowest PUE possible, with some reporting around 1.04 PUE in the ESIG Large Loads Task Force's recent survey. Hot and cold ambient temperatures can affect cooling load requirements and therefore the facility's electric load profile. But high electricity prices and tight grid conditions can cause them to curtail their operations.

FIGURE 21
Variability in Crypto Data Center Demand in ERCOT, January 1, 2025, through July 23, 2025



Source: Morris, S. 2025. "Data Center Inside ERCOT: Observations and Patterns." Slide deck from presentation at ERCOT Large Load Working Group meeting, August 14, 2025. https://www.ercot.com/files/docs/2025/08/11/LLWG_DataCenterObs.pptx.

Importance for the power system: The power use of cryptocurrency mining changes based on known factors such as energy price, cryptocurrency values, and external ambient temperatures. Modelers need to recognize these factors to effectively model operational forecasts of crypto mining.

Load Flexibility

Crypto mining load flexibility is a function of electricity prices relative to cryptocurrency prices. If the cost of energy is too high and mining becomes unprofitable at the current crypto price, the miners will shut down. Most cryptocurrency mining facilities can reduce consumption on request, using software to control demand, although some site load will remain energized to maintain environmental controls and basic infrastructure. For example, some cooling or heating equipment may be kept on, which can require 1% to 2% of total site load.

Some facilities report a preferred limit of 80% load reduction, keeping 20% of load online for auxiliary and parasitic loads, but they can curtail to 95% if needed. Some facilities may need several hours of notice prior to curtailment and/or ramp.

Most cryptocurrency miners in ERCOT participate in ERCOT's Load Acting as a Resource demand response program, curtailing operations to receive a high price for each kWh not consumed during times of grid scarcity. Other crypto miners deliver demand response in the day-ahead time frame, which affects their operational load profile. Most crypto miners self-curtail when the cost of energy exceeds the value of the cryptocurrency that it could produce.

Importance for the power system: Numerous business factors influence a cryptocurrency miner's demand profile in addition to energy and cryptocurrency pricing. These facilities offer highly flexible loads that can provide essential grid reliability services if those services are effectively designed and compensated.

Ramping Capabilities

Cryptocurrency mining facilities can ramp most of their load quickly, taking from 20 seconds to 20 minutes to move operations from 1% to 100% of full load.

Importance for the power system: The fast ramp rates of cryptocurrency miners poses challenges for ensuring adequate balancing and reserve procurement. Batteries are well suited to match the scale and speed of these ramps.

Load Restoration

If a cryptocurrency mining facility trips offline or is curtailed, the load operator will attempt to restore the facility as quickly as possible although the restoration time frame depends on the nature of the trip. The miner's goal is to continue mining as soon as grid conditions are suitable and/or energy market conditions are favorable. Miners with appropriate control software and controls can ramp operations back to full operation quickly. Typically, a cryptocurrency mine can return to normal operations within 20 minutes as it restores its cooling systems.

Importance for the power system: Grid planners and operators need to consider facility restoration times and requirements for load restoration post-fault and develop procedures to manage the restoration impact on balancing reserves procurement.

Electrical Performance

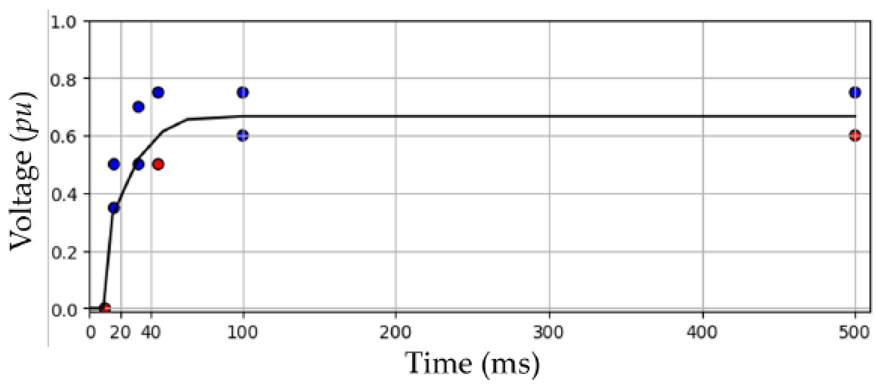
Ride-Through Capability

Cryptocurrency miners are typically designed to the ITIC ride-through capability curve (discussed and illustrated in Figure 3), although, as noted above, since the ITIC curve is not intended to maximize loads' ability to remain grid-connected during grid disturbances, simply meeting ITIC requirements does not necessarily mean that the large load's voltage ride-through capability is sufficient to ensure grid reliability. Experimental data from the S19 Pro Bitcoin miner has shown that the unit surpassed ITIC under-voltage requirements while exhibiting continuous operation from 66% to 130% of the rated voltage. For the zero-voltage ride-through region of the ITIC curve, the miner can only withstand zero voltage for 9.5 milliseconds, compared to 20 milliseconds as required by the ITIC curve (Almubarak et al., 2023). Miners are not designed with special ride-through capabilities outside of the ITIC requirements. Mining rig ride-through capabilities and limitations are driven by the on-board capacitor sizing. To improve on-board ride-through capabilities, miners would need to add large, costly capacitors or batteries to provide voltage response. However, those added costs would likely make this option infeasible.

Figure 22 shows an S19 ASIC miner's ride-through curve in experimental testing. The red dots indicate tests where the miner's power supply tripped and the blue dots indicate tests where the power supply did not trip (Samanta et al., 2024). The experimental validation of the miner's low-voltage ride-through capability was simulated with two faults, one on the miner's premises and one at a nearby solar farm. In the simulation, harmonics from the solar farm negatively interacted with the miner's power supply and led to its tripping. In areas of the grid with a low short-circuit ratio, this behavior could lead to unintended tripping (Samanta et al., 2024). These results show that the miner can meet ITIC requirements. To expand beyond ITIC requirements will require additional hardware at mining facilities or new mining rig designs.

FIGURE 22

S19 ASIC Miner Experimental Ride-Through Curve



This diagram shows an S19 ASIC cryptocurrency miner experimental ride-through curve. The red dots denote where the power supply tripped and the blue dots denote where it did not trip.

Source: A. Samanta, et al., "Electromagnetic Transient Model of Cryptocurrency Mining Loads for Low-Voltage Ride Through Assessment in Transmission Grids," in *2024 IEEE Power & Energy Society General Meeting,* 2024, https://doi.org/10.1109/PESGM51994.2024.10689224.

Importance for the power system: Cryptocurrency miner ride-through capabilities are similar to the performance requirements in the ITIC curve. When performing simulations to asses the facility response to faults, the ITIC curve should be used as a proxy for the ride-through capability in lieu of better information.

Power Quality Requirements and Assessments

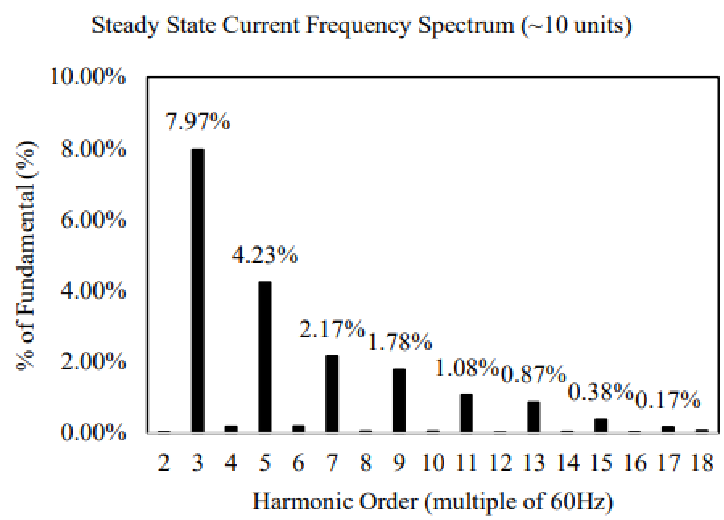
In the Large Loads Task Force's survey and discussions, miners reported the following power quality parameters that must be maintained for optimal operational performance:

- Voltage: Typically ±5% of nominal; three-phase system imbalance should be <2% and sag and swells <1 second typically
- Frequency deviation: ±0.5 Hz or tighter
- Harmonic distortion: Total harmonic distortion (THD) THDv (voltage) of <5% and THDi (current) of <10%

Experimental data from S19 Pro Bitcoin miner has shown that under steady-state conditions, the power factor is between 0.994 and 0.995 leading. During transient conditions the loads show non-linear characteristics. Using data from a large-scale mining facility, researchers measured a voltage THD of 4.3% and current THD of 9.24%, which meets the standards in IEEE 519 (Almubarak et al., 2023).

Experimental data on power quality analysis of AntMiner S9 units has shown that the expected power factor is 0.99 leading (Wheeler et al., 2018). Analysis of harmonics from the miners shows distortion in the triplen harmonics (Figure 23).

FIGURE 23
AntMiner S9 Cryptocurrency Miner's Current Harmonic Spectrum from Experimental Testing



Source:. K. A. Wheeler, et al., "A Power Quality and Load Analysis of a Cryptocurrency Mine," *2018 IEEE Electrical Power and Energy Conference*, Toronto, ON, Canada, 2018, pp. 1-6, https://doi.org/10.1109/EPEC.2018.8598358.

Studies show that with a Delta-Grounded Wye substation transformer, the triplen harmonics will circulate in the delta winding and will not be seen from the transmission system. Here, the delta winding serves as a triplen filter. With no delta winding, the THD is higher and exceeds IEEE Standard 519 harmonic requirements, with the third and fifth harmonics specifically exceeding the requirements and thus requiring filtering. Without the delta winding, the current harmonics circulating back to the transmission grid could reach 9%. A delta winding transformer at the substation would block the third harmonic current to help mitigate this issue; but even with this the harmonic distortion could still exceed allowable limits. Improved power supplies with harmonic filtering and power factor correction would help mitigate these issues.

Several cryptocurrency miners reported that some mining units (e.g., Whatsminers) have tripped offline, likely due to harmonic distortions and voltage fluctuations.

Importance for the power system: Utilities need to conduct power quality and load analysis when interconnecting cryptocurrency mining facilities (Wheeler et al., 2018). A delta winding in the substation transformer would help block the third harmonic current, but additional filter and power factor correction may be needed to ensure that the power quality near a cryptocurrency mine lies within the appropriate limits. Power quality monitoring is necessary at cryptocurrency mining facilites to monitor for harmonics that may impact the power system.

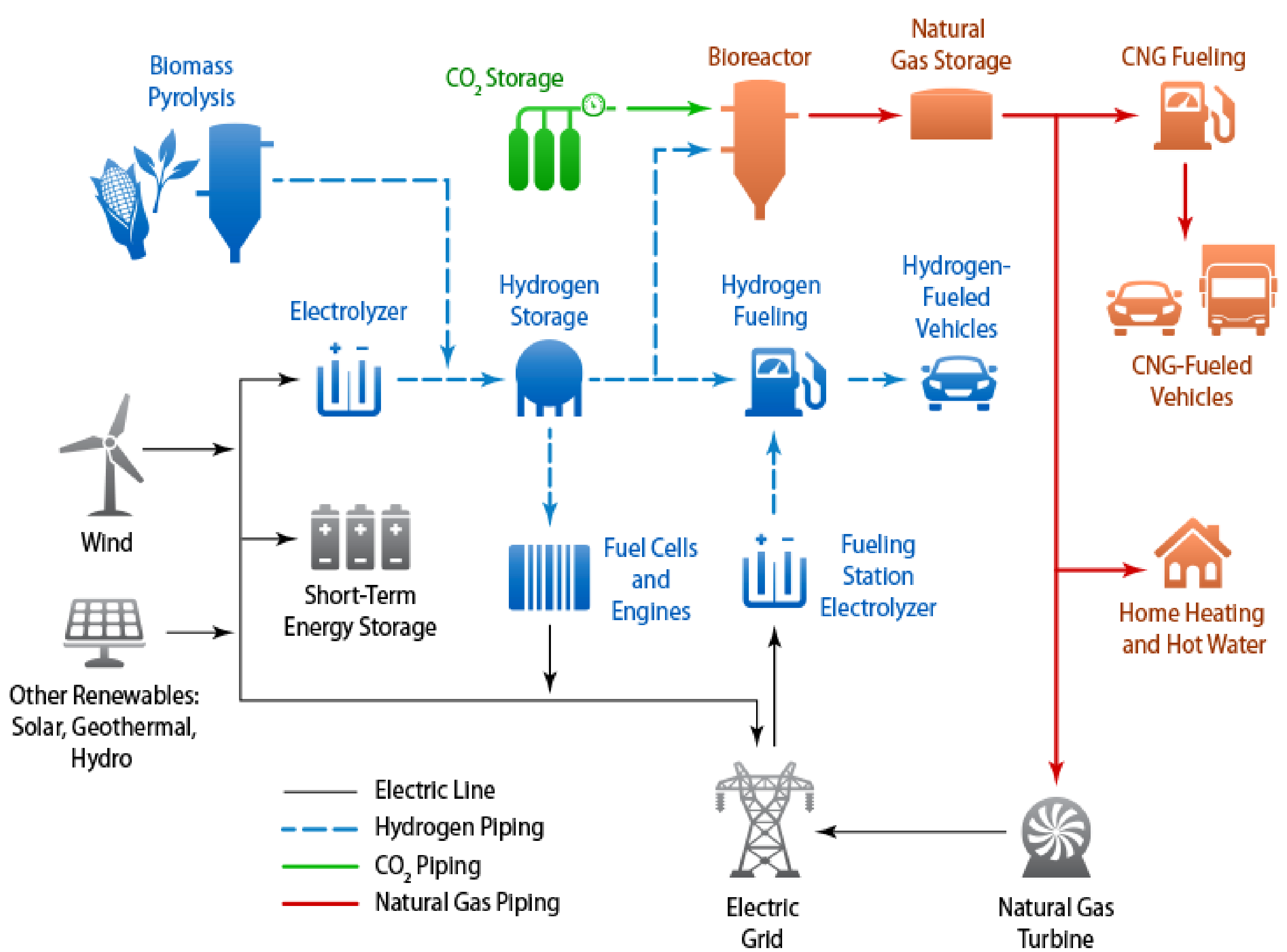
Hydrogen Production

Hydrogen production through electrolysis has gained increasing interest. The modular nature of of electrolyzers allows for a range of sizing options and resulting applications, from small-scale systems powered by renewables to large-scale installations integrated with the electricity grid. The hydrogen produced through electrolysis can serve many applications including long-duration energy storage, fueling stations, industrial processes, residential heating, and in fuel cells to produce electricity.

Figure 24 illustrates several scenarios for producing renewable hydrogen and applications with natural gas and hydrogen-to-grid systems.

FIGURE 24

Various Scenarios for Producing Renewable Hydrogen and Electricity



Source: National Laboratory of the Rockies.

Global hydrogen demand amounted to 97 Mt in 2023, with continual growth since 1975 and more recently showing 3.0% compound annual growth rate over the past decade (IEA, 2024). Hydrogen is predominantly used in refining (43 Mt in 2023) and industrial processes (54 Mt), with minor amounts being used for transportation, buildings, and power generation. Almost all hydrogen today is produced using unabated fossil fuel production methods (natural gas or coal) and the remainder as a by-product in refineries.

Here we focus on industrial-scale electrolytic hydrogen production as a category of large loads. Electrolytic hydrogen production formed less than 1% of global hydrogen production in 2023 (less than 100 kt) but was expected to grow to 17 to 38 Mt by 2030 according to an analysis done in 2023 (IEA, 2023). An indicator of this growth in terms of electric power demand can be seen in electrolyzer capacity reaching 1.4 GW in 2023 (a doubling of 2022 capacity), dominated by expansion in Europe and China. In addition, final investment decisions on 6.5 GW of electrolyzer capacity were reached during 2024 globally. It is not clear at this stage how much of this hydrogen electrolysis capacity will become operational in the short-term (particularly in the U.S., given recent policy changes), but demand growth over the medium term can be expected around 2030 and beyond.

Facility Design

General Design

Electrolytic hydrogen facilities are typically designed to draw their power from a combination of grid power and/or potentially co-located electricity resources, withgreen hydrogen facilities being 100% from renewable resources. The modular nature of electrolyzers allows for various configurations that can be considered during engineering design stages to minimize potential single-point large-load disturbances on the grid. At the same time, most electrolyzer facilities use the grid as back-up power for critical safety loads (e.g., 5% to 10% of nameplate capacity).

Table 3 shows the main components of an electrolyzer facility and their share of the facility’s electrical load.

TABLE 3

Main Components of an Electrolyzer Facility

System/Component	Share of Total Electrical Load	Notes
Electrolyzer stack (electrolysis process)	65–75%	DC conversion losses in power electronics feeding the stack Main H_2 production load
Power converters (e.g., rectifiers, inverters, large power transformers)	5–10%	Large power transformers, which serve banks/modules of electrolyzer stacks and are sized according to the power need Power electronic converters conversion losses + control electronics High efficiency (>95%) but still significant at scale
H_2 compression	10–15%	Depends on output pressure required (some projects have zero compression) High load if hydrogen is compressed to 200–350 bar or more for storage or transport
Water purification and treatment	2–5%	Includes reverse osmosis units, deionization, and pumping Low absolute demand but continuous
Cooling systems	5–10%	Needed for thermal management of stacks, power electronics, and compressor Depends on ambient temperatures and duty cycle
Control systems, lighting, and miscellaneous loads	Less than 1–2%	Supervisory control and data acquisition (SCADA), sensors, emergency systems, HVAC for buildings, and other loads

Source: Energy Systems Integration Group.

Electrolyzer Stack

Most large-scale electrolyzer projects are in the range of 10 MW to 100 MW, and a typical plant is made up of a number of modular stacks. Figure 25 shows an example single-line diagram of an industrial-scale hydrogen production facility (NERC, 2025a). Table 4 lists the typical stack size for each of the three main electrolyzer technologies. More recent facility designs are considering coupling with battery storage, and UPS systems to ensure reliable operation (and for critical safety loads) as well compliance with utility specific interconnection requirements.

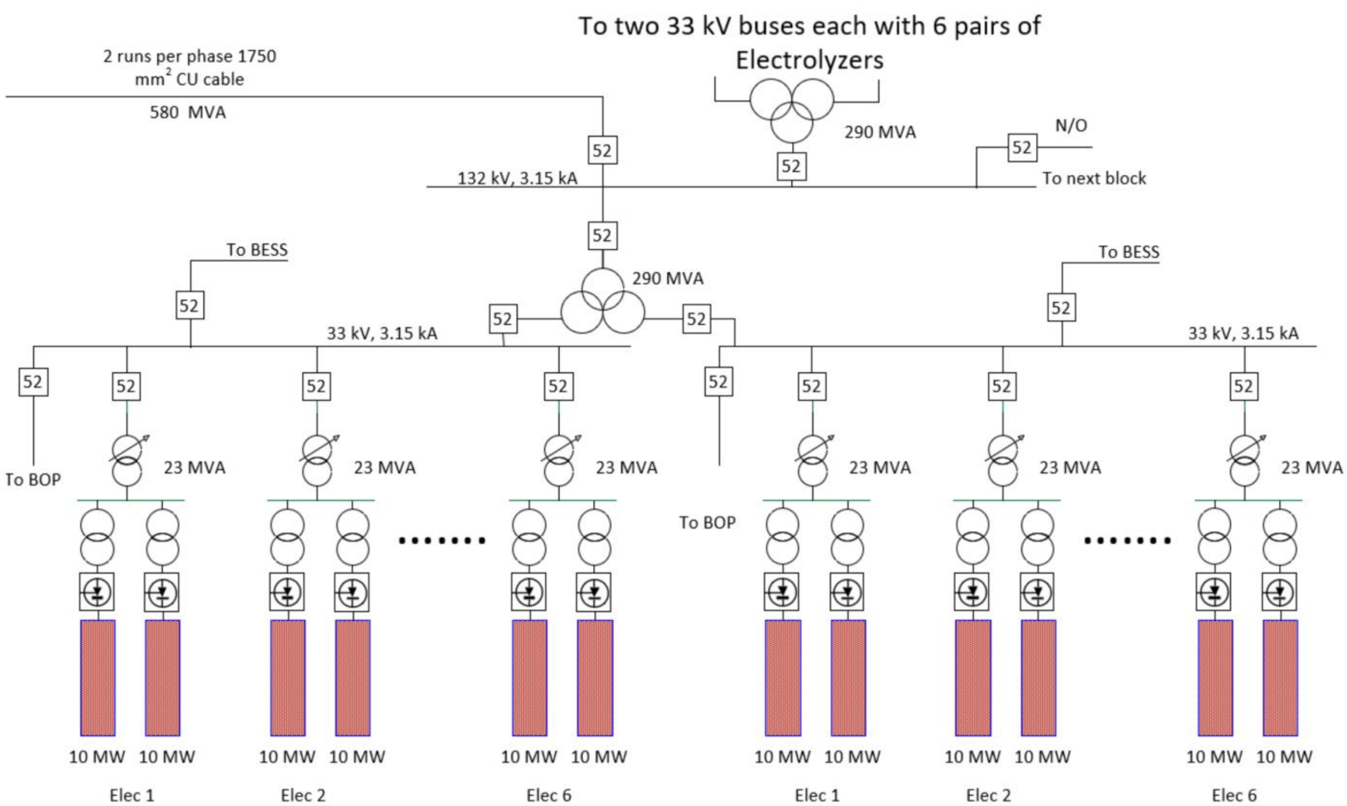
The main types of electrolyzers are:

- **Alkaline electrolyzers (AEL):** This is the most mature of electrolyzer technologies at large scale. These employ an alkaline electrolyte solution, like potassium hydroxide, and nickel electrodes, operating at lower temperatures with a long history of commercial applications. They are typically preferred to operate at a stable base-load.

- **Proton exchange membrane (or polymer electrolyte membrane) (PEM) electrolyzers:** PEM electrolyzers use precious metal catalysts and a solid polymer membrane to separate gases, operating at higher temperatures and pressures for improved efficiency and rapid response.
- **Solid oxide electrolytic cell (SOEC) electrolyzers**: SOEC electrolyzers rely on solid ceramic electrolytes functioning at high temperatures with steam as an input. SOECs are aslo capable of converting carbon dioxide and water vapor directly into syngas, showing promise for carbon capture applications.
- **Anion exchange membrane (AEM) electrolyzers**: Still in developmental stages, the AEM technology aims to combine the need for increased ramp rates and high pressure advantages of PEM electrolyzers with lower capital cost of AEL electrolyzers.

FIGURE 25

Example Single-Line Diagram of an Industrial-Scale Hydrogen-Production Facility



Notes: BESS = battery energy storage system; BOP = balance of plant.

Source: EPRI, reproduced in North American Electric Reliability Corporation, "Characteristics and Risks of Emerging Large Loads," Large Loads Task Force White Paper (2025), https://www.nerc.com/globalassets/who-we-are/standing-committees/rstc/whitepaper-characteristics-and-risks-of-emerging-large-loads.pdf.

TABLE 4
Typical Stack Size for Electrolyzer Technologies

Technology	Stack size (MW)
PEM (Proton Exchange Membrane)	0.5–2.5 MW
Alkaline (AEL)	1–5 MW
SOEC (Solid Oxide)	0.1–1 MW

Source: ENERTRAG.

Rectifiers

Rectifiers can be categorized into two main technologies, thyristor based silicon-controlled rectifiers (SCRs) and insulated gate bipolar transistor (IGBT) rectifiers, both of which convert AC electricity to DC electricity for use by electrolyzers. Large-scale electrolytic hydrogen production typically uses SCR rectifiers (diode or phase-controlled, configured as 6-pulse, 12-pulse or 24-pulse configurations) due to the capabilities of the underlying power electronics in SCR rectifiers and their low tolerance to fault conditions. For electrolyzers that need more than 10 MW, IGBT rectifiers cannot be used because they are current-limited. Twenty-four-pulse thyristor-based rectifiers are being considered to reduce harmonics emissions or the need for mitigating equipment; when 24-pulse or 48-pulse thyristor-based rectifiers are not used, the electrolyzer plant can be supplemented with mitigating equipment like distribution static synchronous compensators (DSTATCOMs). Smaller PEM electrolyzers are beginning to use alternative power electronics (IGBTs) to resolve power quality issues. However, the voltage sensitivity of these non-linear loads can worsen voltage dips under certain conditions.

IGBT recifiers are gaining prominence due to improved power quality (harmonics performance, power factor) as well as improved performance in weak grid conditions (Zeng et al., 2026; Tavanaee, 2025). However, their switching losses can be higher, and they have the potential to cause electromagnetic interference. IGBT rectifier costs can also be higher relative to SCR rectifiers. SCR rectifiers are often considered more robust and cost-effective for high-power applications, but their performance becomes challenging due to high reactive power consumption and the potential for commutation failures (self-commmutating power electronics).

Battery storage and UPS systems can be deployed to provide back-up for safety-critical loads and facilitate safe plant shutdown. Post-outage preparations require operators to reset interlocks, re-energize critical equipment, and perform standard start-up. Automatic reconnection to the grid typically occurs at the high-voltage/medium-voltage level at the high-voltage winding of large power transformers.

Rectifier Transformers

Specialized large power transformers called rectifier transformers can be packaged with rectifiers considering their application-specific needs (providing input AC power to a rectifier front-end). These transformers step down voltage from bulk grid supply (where the electrolyzer is grid-connected or interfaced with any co-located power sources like wind, solar, battery storage, and/or thermal power generation). These transformers can also be arranged as pairs of transformers supplying rectifier converters to create a 12-pulse rectifier system. The transformers have different vector groups in order to minimize harmonic emissions to the grid (specifically 5^{th} and 7^{th} order harmonics) and reduce DC current ripple being supplied to electrolyzers.

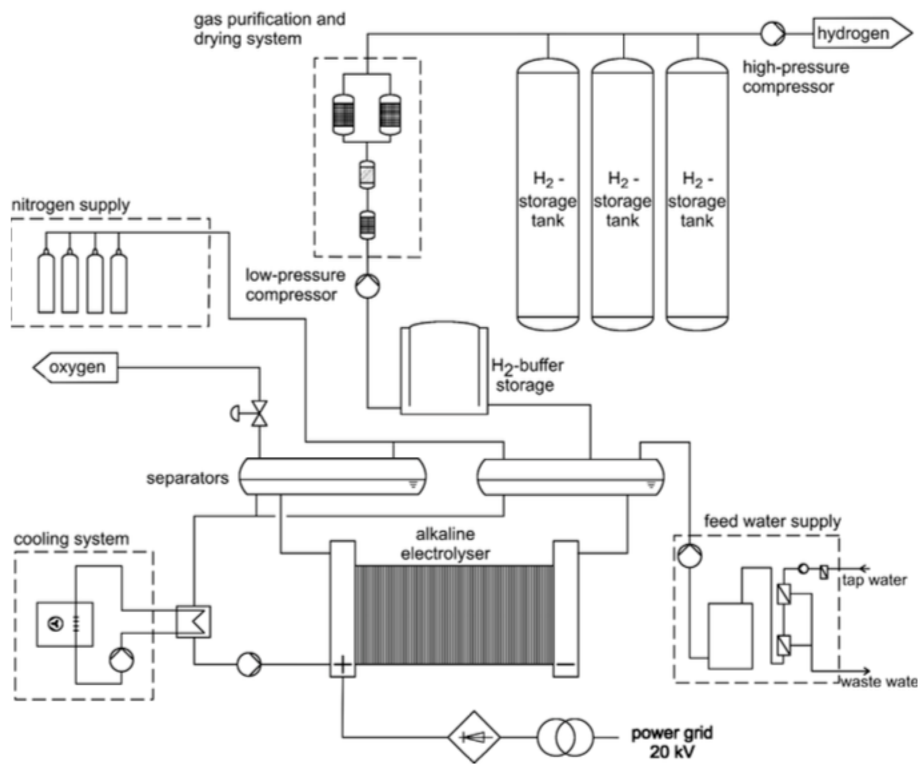
Balance of Plant

The main components of an industrial-scale AEL electrolyzer is shown in Figure 26 (Fischer et al., 2016). Important components, and their power demand requirements, are:

- **Water purification:** Water purification is a key preparatory process for hydrogen electrolyzers, and power requirements are typically 2% to 5% of total electrolyzer plant load. Purification steps include pre-filtration to remove particulates, reverse osmosis or ion-exchange, deionization, and ultraviolet (UV) sterilization in some cases. PEM and SOEC electrolyzers are more sensitive to water purity than the more durable AEL electrolyzers. The loads related to water purification are dominated by heating for steam creation (SOEC electrolyzers only), pumps and motor loads (sometimes interfaced through VSDs), followed by smaller rectifier loads for deionization and AC-coupled loads like UV lamps for sterilization.
- **Compression:** Hydrogen compression is essential to enable storage, transportation, and integration with end-use applications, and power requirements are typically 10% to 15% of total electrolyzer plant load. Pressure levels vary depending on electrolyzer type and potential end-use application—typical storage in local pressure vessels is 20 bar while further compression in pipelines or long-term storage is 100 to 200 bar. Generally, AEL electrolyzers produce hydrogen at atmospheric pressure and therefore require more compression equipment (compressing to 10 to 40 bar). PEM electrolyzers operate at 10 to 30 bar and require relatively less compression equipment for post hydrogen production compression, while SOEC electrolyzers operate at above atmospheric levels depending on the application. Compressor loads are mostly motor-driven and often interfaced through VFDs, with other smaller loads like auxillary cooling, control and instrumentation, and power quality equipment making up the remainder of compressor loads.
- **Cooling systems:** Cooling systems for electrolyzers mostly focus on cooling the electrolyzer stack (and also the electrolyte, for AEL electrolyzers). They typically use cooling loops with water or coolant (sometimes air cooling) circulated through heat exchangers, chillers, and pumps, with typical power requirements of 5% to 10% of total electrolyzer plant load. SOEC electrolyzers have very different temperature control equipment and typically have lower cooling needs because they naturally operate at higher temperatures (with strict temperature control) relative to AEL and PEM electrolyzers. Electrical loads typically comprise pumps and blowers (motor loads), with a smaller role for auxilliary equipment providing control logic and telemetry.

FIGURE 26

Main Components of Industrial-Scale Alkaline Electrolyzer



Source: U. R. Fischer, et al. 2016. "Hydrogen Hybrid Power Plant in Prenzlau, Brandenburg," in *Hydrogen Science and Engineering: Materials, Processes, Systems, and Technology*, edited by D. Stolten and B. Emonts. Wiley-VCH. https://onlinelibrary.wiley.com/doi/10.1002/9783527674268.ch44.

Operational Characteristics

Load Profile

Electrolyzer facilities are typically operated at a relatively constant power setpoint to minimize the number of starts and stops that the equipment experiences. But while operation at baseload is typical, it can be adjusted based on on-site supply resource availability (wind/solar supply variability), wholesale electricity market prices, and/or hydrogen demand, with storage and supply-side flexibility potentially being used to buffer electrolyzer operational and power use fluctuations. Load profiles of electrolyzer plants are expected to change over time due to hydrogen contractual obligations, prices of wholesale grid power, the availability of supply resources (particularly relevant for green hydrogen projects), and electrolyzer stack aging. Load profiles also depend on manufacturer specifications, technology, and maintenance protocols of electrolyzers.

The critical infrastructure load component of an electroylzer facilites demand is usually quite low (typically less than 5%) and includes protection of the auxiliary equipment (e.g., controls, progammable logic controllers

(PLCs), safety valves, and cooling systems) necessary to ensure safe shutdown and prevent hazards in the event of power loss. Facility design will incorporate protective measures for controlled shutdowns during disruptions.

Load Flexibility

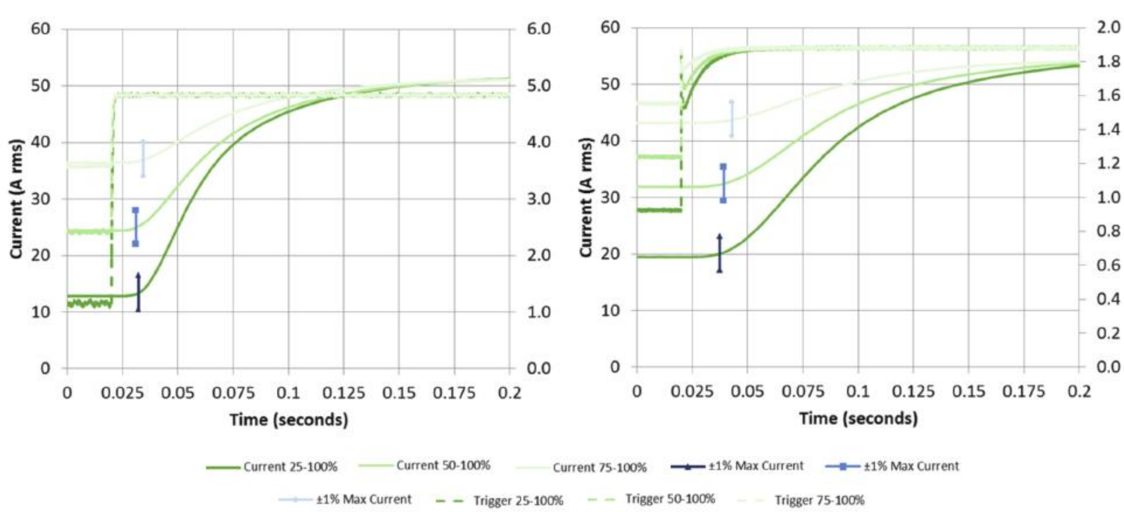
Green hydrogen facilities are expected to be capable of reducing or increasing electricity consumption upon request and therefore can be considered a controllable or flexible load. This has mostly been driven by previously described drivers—hydrogen demand/contractual obligations, supply resource availability, and wholesale market prices. The electrochemical process has the potential to restrict the ability of electrolyzers to adjust or shift electrical load too rapidly, at speeds faster than the grid can accommodate, particularly for AEL electrolyzers, which have liquid membranes and hence potentially slower response times than PEM, SOEC, and AEM electrolyzers.

Based on the operational characteristics discussed above, electrolyzers could be used to provide various ancillary services to the grid. The technical ability of AEL and PEM electrolyzer loads to provide ancillary services was demonstrated by Eichman, Harrison, and Peters (2014), with illustrations of ramp time capabilities (slew-rate) in Figure 27 and demand-side response as primary frequency control in Figure 28. Equivalent inertial response is shown in Figure 29 (Jain et al., 2023), the provision of both active power (primary and secondary frequency control) as well as reactive power (voltage support) shown in Figure 31 through a hardware-in-the-loop (HIL) environment (Nguyen et al., 2022), and second-to-second (load-following) control shown in Figure 31 for an operational electrolyzer plant in Germany (Fischer et al., 2016).

Whether these ancillary services are compensated or are required as part grid code or interconnection standards, the technical capability exists and the appropriate control hierarchy can be established to consider relevant grid code requirements and expected grid events.

FIGURE 27

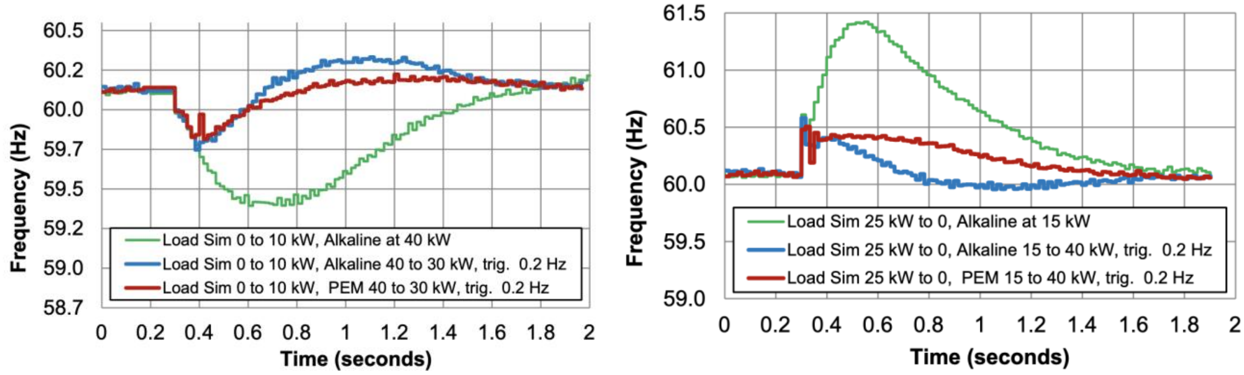
Ramp-Up Demonstration for Small-Scale PEM (Left) and Alkaline (Right) Electrolyzers



Source: J. Eichman, K. Harrison, and M. Peters, *Novel Electrolyzer Applications: Providing More Than Just Hydrogen* (National Renewable Energy Laboratory, 2014), https://doi.org/10.2172/1159377.

FIGURE 28

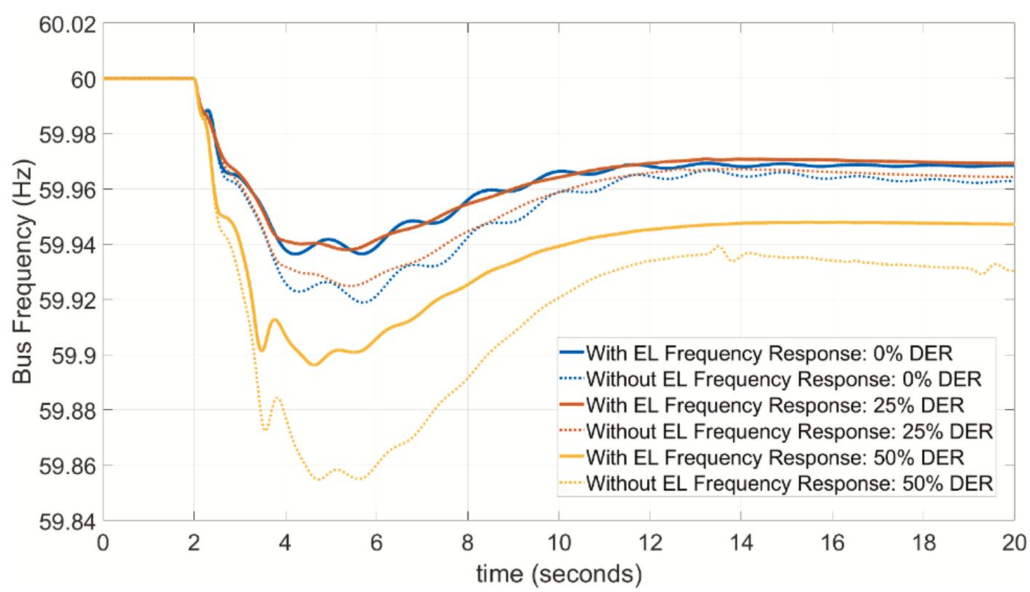
Demonstration of Frequency Response Up (Left) and Down (Right) for Small-Scale PEM and Alkaline Electrolyzers



Source: J. Eichman, K. Harrison, and M. Peters, *Novel Electrolyzer Applications: Providing More Than Just Hydrogen* (National Renewable Energy Laboratory, 2014), https://doi.org/10.2172/1159377.

FIGURE 29

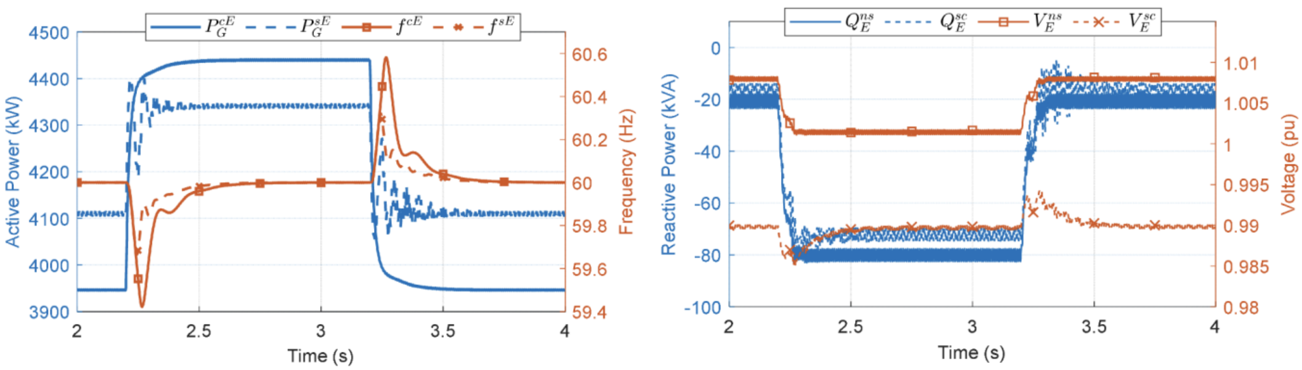
System Frequency Response Under Varying Levels of Distributed Energy Resources with Small-Scale PEM and Alkaline Electrolyzers' Frequency Response Enabled or Disabled



Source: R. Jain, et al., "Grid Ancillary Services Using Electrolyzer-Based Power-to-Gas Systems with Increasing Renewable Penetration," *e-Prime - Advances in Electrical Engineering, Electronics and Energy* 6 (2023): 100308, https://doi.org/10.1016/j.prime.2023.100308.

FIGURE 30

Demonstration of Active Power (Primary and Secondary Frequency Control) and Reactive Power (Voltage Support) in a Hardware-in-the-Loop Environment

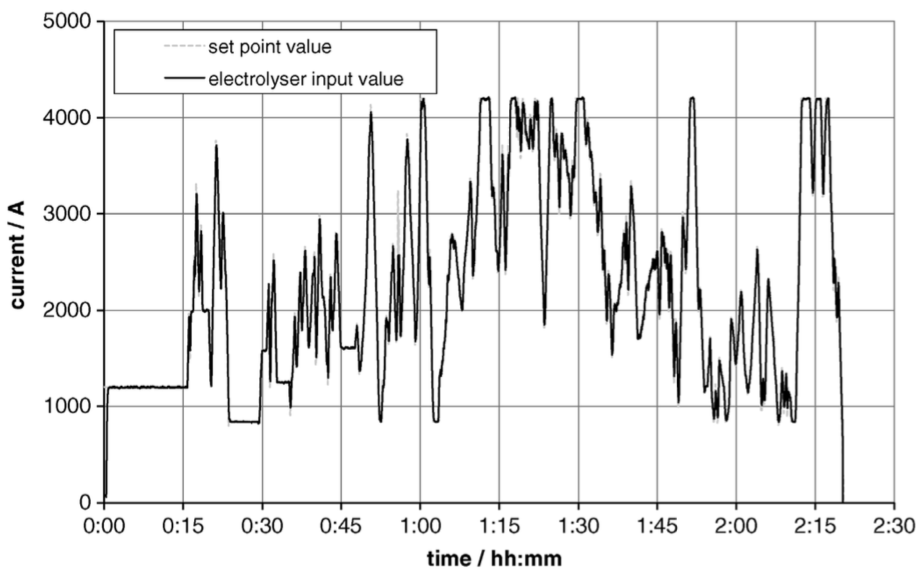


(Left) Active power and frequency response after load switching (2.2 s and 3.2 s) with constant electrolyzer and supporting electrolyzer under secondary control. (Right) Reactive power and voltage responses with constant electrolyzer and supporting electrolyzer under secondary control.

Source: B. L. H. Nguyen et al., "Hierarchical Control of Grid-Connected Hydrogen Electrolyzer Providing Grid Services," in *2022 IEEE 1st Industrial Electronics Society Annual On-Line Conference (ONCON)*, 2022, https://doi.org/10.1109/ONCON56984.2022.10126806.

FIGURE 31

Example of a Dynamic Response (Setpoint-Following) of a MW-Scale Alkaline Electrolyzer



Source: U. R. Fischer, et al. 2016. "Hydrogen Hybrid Power Plant in Prenzlau, Brandenburg," in *Hydrogen Science and Engineering: Materials, Processes, Systems, and Technology*, edited by D. Stolten and B. Emonts. Wiley-VCH. https://onlinelibrary.wiley.com/doi/10.1002/9783527674268.ch44.

Ramping Capabilities

The ramp rate of PEM electrolyzers is typically 1% to 20% per second of DC current; depending on the manufacturer, this speed could be down to 5% to 12% per second). Other types of electrolyzers, such as alkaline electrolyzers, have ramp rates of 0.5% to 2% per minute of DC current.

Facility Start-Up

The time it takes to return to normal operation after a plant is curtailed or trips offline depends on the plant's condition after a trip event. Start-up times are classified as cold start and warm start:

- **Cold start** represents a start with the unit having been off long enough to lose benefit from the energy stored in the electrolyzer (e.g., thermal energy for lubricants and coolants). In a cold start, it is possible to ramp from 0 to 100% in as little as 30 to 60 minutes for alkaline electrolyzers and 3 to 15 minutes for PEM electrolyzers, though ranges of 1 to 2 hours are more common. Restarting from a complete outage/black start-up (including safety-critical equipment) could require a longer recovery time.
- **Warm start** represents a start with the unit having been off or in standby for a short time, allowing it to benefit from energy stored in the electrolyzer. In a warm start, a hydrogen facility is typically capable of ramping from 0% to 100% nameplate load in about 5 to 30 minutes. Once the electrolyzer has begun operating, 10% per second ramp-up is possible.

Minimum Turndown

The minimum turndown is the lowest operating point for electrolyzer power consumption, after which the device must turn off. A typical hydrogen production facility cannot reduce production below a pre-specified value. Only AEL electrolyzers have sufficiently binding constraints that require operation above a minimum turndown (typically 5% to 25% depending on facility design) in order to avoid safety risks (gas cross-over of hydrogen/oxygen that could be explosive).

Electrical Performance

Voltage and Frequency Ride-Through

The voltage and frequency ride-through capability of an electrolyzer facility is normally limited by the electrochemical process and its dynamics within and behind the electrolyzer stacks. Example voltage and frequency ride-through settings of a typical electrolyzer facility are shown in Table 5 (but these are by no means prescriptive). Ride-through limitations for electrolyzers are driven mostly by the electrochemical process and its dynamics within and behind the electrolyzer stacks (not by thyristor- or IGBT-based converter capabilities). Supplementary equipment has the potential to support ride-through capabilities but is typically applied on a project-specific basis (with requisite cost trade-offs on the overall cost of production).

Following a fault and return to nominal voltage grid voltage and frequency, the post-fault recovery ramp rate for electrolyzers is typically 10% to 15%/s.

TABLE 5
Typical Voltage (top) and Frequency (bottom) Ride-Through for an Electrolyzer

U [p.u.]	Time [s]
0–0.7	0.3
0.7–0.9	2.25
0.9–1.1	Continuous
1.1–1.2	3
1.2–1.3	0.1
>1.3	0

f [Hz]	Time (s)
<47.5	Trip
47.5–48.5	1800
48.5–49	1800
49–51	Continuous
51–51.5	1800
51.5–52.5	10 seconds
>52.5	Instantaneous trip

Source: Energy Systems Integration Group.

Other Characteristics

Hydrogen facilities also have the following characteristics, which determine how the facility responds during faults, and can influence power quality in the local power grid.

- **Electrical protection:** Under/over-voltage and under/over-frequency protection is established within the plant to prevent damage of equipment in response to unexpected grid disruptions or voltage irregularities.
- **Phase angle change:** Electrolyzer operation is typically insensitive to phase angle changes up to a certain value (e.g., up to +/- 30 deg).
- **Power quality**: Intentional rectifier converter design is applied to reduce emitted harmonics (e.g., 24-pulse thyristo-based rectifiers). Additional mitigation equipment is sometimes needed to meet harmonics obligations (passive or active depending on requirements and network conditions). The deployment of IGBT rectifier converters has the potential to support resolving power quality issues (usage is established through design to ensure controller stability). Voltage unbalance should be kept

within limits to guarantee optimal performance of rectifier transformer(s), i.e., to not cause additional losses/heating. Available short-circuit power at the POI should be kept above a certain level; this is highly applicable for LCC converters, and specific short-circuit power needs are technology- and vendor-specific.

Appendix: Overview of Large Load Categories and Configurations

This section summarizes the various service configurations for large loads, broken down into four categories. The first category represents data centers, which are the most significant area of concern addressed by the ESIG Large Loads Task Force. The second category is large industrial loads, which have become more prevalent with increasing manufacturing and mining activities. The third is the production, storage, and transportation of hydrogen, and the fourth is electric vehicle (EV) charging demand.

Category 1: Data Centers and Computational Load

TABLE A-1
Data Centers

Component	Power-Consuming Devices	Function / Service	Uncertainty Level	Flexibility Potential	Notes	Importance to the Grid
IT-related equipment	Servers (CPU, GPU, application-specific integrated circuit (ASIC), field-programmable gate array (FPGA)) Storage arrays (solid-state drive, hard disk drive) Networking hardware (routers, switches)	Data processing, storage, and transmission Core computing function (e.g., AI model training, search queries, video processing)	Low Highly predictable workloads with consistent 24/7 operation, especially in hyperscale data centers. Job scheduling and service level agreements minimize variability.	Low Interrupting load directly leads to service degradation or outages	Highest energy consumption share (~45–55%) Sensitive to latency and thermal fluctuations Operates nearly 24/7 for hyperscale workloads	Data centers with CPU-based compute generally have a more stable load profile due to both their workloads and sequential processing design. In comparison, today's more common GPU-based compute has a more variable load profile due to its workloads and parallel processing capabilities.
Power distribution system	Uninterruptible power supply (UPS) Power distribution units	Ensures power quality, voltage regulation, back-up in outages	Moderate While the system is designed for reliability, activation of UPS/battery	Moderate Some UPS and batteries can be configured for grid support	Normally passive but can be used for grid ancillary services	The design of power distribution systems has numerous impacts on the power system. Power conversion equipment design determines the capabilities and limitations of the equipment to withstand fault-induced

Component	Power-Consuming Devices	Function / Service	Uncertainty Level	Flexibility Potential	Notes	Importance to the Grid
	Inverters and rectifiers Back-up generators (diesel, gas) Battery banks (valve regulated lead-acid battery (VRLA), Li-ion)	Transient power support	systems can vary with grid conditions, maintenance cycles, or integration with external signals.	(e.g., frequency response, fast dispatch)	Flexibility depends on integration with grid coordination systems	low voltages. Increasing inverter overload capability may help mitigate fault-induced voltage problems and improve fault ride-through. To properly simulate the data center, modelers will need to create positive-sequence phasor domain (PSPD) and electromagnetic transient (EMT) models for the new architectures including the sidecar, SST, and other power conversion equipment. Power conversion equipment could become a source of troublesome harmonics that may need study and mitigation.
Cooling systems	Computer room air conditioning units Chillers and cooling towers Pumps and fans Liquid immersion or rear-door cooling (for high-performance computing or AI)	Heat removal to maintain operational temperature of IT equipment	Moderate Cooling load varies with ambient conditions, IT workload intensity, and system design (e.g., economizer mode).	High (depending on design) Thermal inertia allows short-term curtailment or duty cycling	Typically accounts for ~25–35% of energy use Cooling load can be shifted in time (e.g., pre-cooling, thermal storage) Integration with outside air economizers adds more flexibility	Data centers with cooling systems driven by variable frequency drives may be susceptible to ride-through issues with their VFDs and IT equipment. Cooling system failure would trigger a data center shutdown if redundant cooling systems fail. The cooling load affects the data center PUE and overall energy consumption. Understanding the cooling system will enable more accurate forecasting of short-term responses to grid events and longer-term load profiles.

Notes: CPU = central processing unit; GPU = graphics processing unit; IT = information technology; PUE = power usage effectiveness. Note: IT-system power consumption is usually coupled with the cooling system.

Source: Energy Systems Integration Group.

TABLE A-2
AI Training and Inference Facilities

Component	Power-Consuming Devices	Function / Service	Uncertainty Level	Flexibility Potential	Notes	Importance to the Grid
Compute infrastructure	AI accelerators: GPUs (e.g., NVIDIA A100/H100), TPUs, ASICs CPUs (for orchestration and IO) High-speed memory (HBM, DDR)	Executes AI model training or inference tasks (e.g., LLMs, vision models)	Low (training) Training jobs are long-duration, scheduled, and often bound by deadlines and compute reservation windows. Moderate (inference) Inference workloads—especially batch or offline—can have variable timing and be rescheduled within service level agreements.	Low (training) Moderate (inference)	AI training is time-intensive and sensitive to interruptions Inference (esp. batch jobs) can be deferred or rescheduled in some frameworks	Data centers with CPU-based compute generally have a more stable load profile due to both their workloads and sequential processing design. In comparison, today's more common GPU-based compute has a more variable load profile due to its workloads and parallel processing capabilities.
Power conversion and distribution	UPS and battery banks PDUs Power converters and transformers	Power stability, voltage/frequency control, back-up power	Moderate	Moderate Can offer grid services (e.g., fast frequency response, back up to grid)	High-quality power is essential for AI workloads Li-ion batteries often oversized and underutilized	The design of power distribution systems has numerous impacts on the power system. Power conversion equipment design determines the capabilities and limitations of the equipment to withstand fault-induced low voltages. Increasing inverter overload capability may help mitigate fault-induced voltage problems and improve fault ride-through.

Component	Power-Consuming Devices	Function / Service	Uncertainty Level	Flexibility Potential	Notes	Importance to the Grid
						To properly simulate the data center, modelers will need to create PSPD and EMT models for the new architectures including the sidecar, SST, and other power conversion equipment. Power conversion equipment could become a source of troublesome harmonics that may need study and mitigation
Cooling systems	Direct liquid cooling (DLC) Rear-door heat exchangers Precision chillers Coolant pumps and fans	Dissipates heat from densely packed AI compute units	Moderate	High With thermal buffering, cooling load can be shifted temporarily	AI clusters generate higher thermal density than standard data centers Facilities often designed with high-efficiency cooling, enabling pre-cooling or load cycling	Data centers with cooling systems driven by VFDs may be susceptible to ride-through issues with both their VFDs and IT equipment. Cooling system failure would trigger a shutdown of the data center if redundant cooling systems fail. Data center cooling system load affects the data center PUE and its overall energy consumption, so understanding the nature of the cooling system will enable more accurate forecasting of both short-term responses to grid events and longer-term load profiles.

Notes: AI = artificial intelligence; ASIC = application-specific integrated circuit; CPU = central processing unit; DDR = double data rate; GPU = graphics processing unit; HBM = high bandwidth memory; IO = input/output; LLM = large language model; PDU = power distribution unit; TPU = tensor processing unit; UPS = uninterruptible power supply.

Source: Energy Systems Integration Group.

TABLE A-3

Cryptocurrency Mining Facilities

Component	Power-Consuming Devices	Function / Service	Uncertainty Level	Flexibility Potential	Notes	Importance to the Grid
Mining hardware	Application-specific integrated circuit (ASIC) miners (e.g., Bitmain Antminer, WhatsMiner) Power-hungry GPUs (legacy) Hash boards and controllers	Solves cryptographic hashes to validate blockchain transactions	High Mining load can fluctuate based on market prices, pool participation, and scheduled profitability windows. Operators can adjust compute intensity rapidly.	High Loads can be ramped down almost instantly	Crypto mining is inherently interruptible Profitability-driven operations allow participation in demand response or price-based curtailment	Cyptocurrency mining hardware will inform load profiles and power system models.
Power supply and conditioning	Switchgear Transformers and rectifiers Back-up diesel gensets (rare)	Conditions power and ensures safe delivery to rigs	Moderate	Moderate (if BESS/UPS integrated)	Most systems are optimized for uptime, but rarely grid-interactive Growing interest in "grid-balancing mining farms"	Cyptocurrency miner models must represent the component mining rigs accurately to capture the facility's operations and electricity use effectively.
Cooling systems	Air-cooled fans (primary) Evaporative cooling Immersion cooling systems	Prevents thermal throttling and equipment degradation	Moderate	High Load can be cycled with minimal impact	Lower cooling requirements than AI/data centers Advanced farms use immersion cooling to reduce noise and increase efficiency	In addition to IT equipment failing to ride through a grid disturbance, cooling systems driven by VFDs are also susceptible to ride-through failures. If the cooling system fails to ride through or shuts down for another reason, the entire cryptocurrency mining facility will shut down, exacerbating the impact of the intial disturbance. VFDs in cooling systems at cryptocurrency mining facilities should be reviewed and set to ensure they do not trip off during normal faults.

Notes: BESS= battery energy storage system; GPU = graphics processing unit; UPS = uninterruptible power supply; VFD = variable frequency drive.

Source: Energy Systems Integration Group.

Category 2: Industrial Load Types

The loads detailed here in the industrial load category may not capture all existing and potential future industrial load types. Most industrial demand characteristics are well understood; therefore, uncertainty levels are not included for simplicity.

TABLE A-4

Mining and Mineral Processing

Component	Power-Consuming Devices	Function / Service	Flexibility Potential	Remarks
Crushing and grinding	Jaw crushers Ball mills Semi-autogenous grinding (SAG) mills Conveyor belts	Reduces ore to manageable size for further processing	Moderate (batch or staged processing)	Energy-intensive, large motors with controllable duty cycles
Material transport	Belt conveyors Bucket elevators Pumps	Moves slurry, crushed rock, or finished materials	High (interruptible if storage buffers exist)	May be interruptible depending on storage design
Refining and separation	Flotation cells Magnetic separators Electrochemical cells	Extracts valuable minerals/metals	Low for electrochemical Moderate for mechanical	Electrochemical steps (e.g., leaching) often require consistent operation
Ventilation and dewatering	Large fans Dewatering pumps	Safety and environmental control	High (short-term)	Can be managed during off-peak hours

Source: Energy Systems Integration Group.

TABLE A-5

Metals and Heavy Manufacturing

Component	Power-Consuming Devices	Function / Service	Flexibility Potential	Remarks
Melting and forming	Electric arc furnaces (EAF) Induction furnaces Rolling mills	Melting scrap or raw ore and shaping into products	Moderate to High for batch operations	EAFs are known flexible loads used in demand response programs
Process equipment	Casting machines Extruders Lathes and CNCs	Converts melted metals into usable products	Low to Moderate	Critical for product quality; delays can disrupt entire production
Compressed air systems	Compressors Blowers	Powers pneumatic tools, cleaning, control	High (with tank storage)	Can be shifted to off-peak hours; significant in some facilities
Material handling	Cranes Conveyors Loaders	Moves raw and finished goods	High (if buffers present)	Not mission-critical in continuous processes

Notes: CNC = computer numerical control.

Source: Energy Systems Integration Group.

TABLE A-6

Semiconductor and Electronics Manufacturing

Component	Power-Consuming Devices	Function / Service	Flexibility Potential	Remarks
Cleanroom HVAC	HEPA filters Laminar flow fans Chillers	Maintains sterile environment and air purity	Moderate (limited cycling possible)	Critical but may have thermal/humidity tolerances
Fabrication tools	Photolithography machines Etchers Ion implanters	Patterning and doping of silicon wafers	Low	Precision processes requiring strict power quality
Process cooling	Liquid chillers Water circulation systems	Removes heat from process tools	High (with thermal buffering)	Usually backed with UPS or BESS
Power supply and conditioning	UPS, isolation transformers Power quality monitors	Stable voltage and clean power for delicate electronics	Moderate	Power stability critical, but some BESS can support grid

Notes: BESS= battery energy storage system; HEPA = high efficiency particulate air; HVAC = heating, ventilation, and air conditioning; UPS = uninterruptible power supply.

Source: Energy Systems Integration Group.

TABLE A-7

Chemical and Petrochemical Processing

Component	Power-Consuming Devices	Function / Service	Flexibility Potential	Remarks
Reaction units	Electrolyzers (e.g., for ammonia, hydrogen) Reactors (catalytic, thermal)	Drives chemical transformation of feedstocks	Moderate (depending on batch vs. continuous)	Batch reactors more flexible; continuous systems less so
Pumps and compressors	High-head pumps Reciprocating compressors	Moves gas/liquids across plant stages	High (if flow buffers or storage tanks exist)	Can be managed dynamically depending on process flow
Utilities	Boilers Steam generation HVAC systems	Provides heating, cooling, and pressure	High (especially with thermal storage)	Typically, easier to shift load with proper system design

Notes: HVAC = heating, ventilation, and air conditioning.

Source: Energy Systems Integration Group.

TABLE A-8

Oil and Gas Production

Component	Power-Consuming Devices	Function / Service	Flexibility Potential	Remarks
Drilling and pumping	Submersible pumps Beam pumps Compressor stations	Extract and transport oil/gas	Moderate for some well types	Some flexibility depending on well design and flow constraints
Gas processing	Refrigeration units Fractionators Dehydrators	Separates gas into usable components	Low to moderate	Often continuous-flow process with limited curtailment capacity
LNG operations	Cryogenic liquefiers Storage systems	Converts natural gas to LNG for transport	Low	High safety and thermal sensitivity; minimal flexibility
Electrification of equipment	Electrified compressors, drills, rigs	Transition from diesel/natural gas to grid power	Moderate	A growing source of load; flexibility depends on system design and contractual commitments

Notes: LNG = liquified natural gas.

Source: Energy Systems Integration Group.

Category 3: Hydrogen Production Facilities

TABLE A-9

Hydrogen Production Facilities

Component	Power-Consuming Devices	Function / Service	Flexibility Potential	Remarks
Electrolyzer stack	Proton exchange membrane (PEM) electrolyzers Alkaline electrolyzers Solid oxide electrolytic cell (SOEC) electrolyzers	Splits water into hydrogen and oxygen using electricity	Moderate (especially PEM)	Most electrolysis systems can ramp up/down quickly Flexibility varies by technology (PEM > Alkaline > SOEC)
Hydrogen compression and storage	Mechanical compressors Electric pumps Cryogenic coolers	Compresses and stores hydrogen in tanks, pipelines, or liquefaction systems	Moderate (when storage buffer is available)	Can shift energy use based on electricity prices or grid needs Storage enables temporal decoupling of production and delivery

Source: Energy Systems Integration Group.

Category 4: Transportation Load

Electric vehicle (EV) loads can be categorized under industrial loads, following the classification in the NERC Large Load Task Force White Paper, "Characteristics and Risks of Emerging Large Loads" (NERC, 2025a).

TABLE A-10

EV Charging Facilities

Component	Power-Consuming Devices	Function / Service	Uncertainty Level	Flexibility Potential	Remarks
DC fast chargers (depot)	50–350 kW DCFC units per port for small vehicles Up to 1MW+ DCFC units per port for large vehicles combined charging systems, MCS (for HDVs)	Charges lightmediumand heavy-duty EVs quickly	High Usage depends on fleet schedules, travel patterns, and charging behavior.	Moderate High capability for demand response but the point of fast chargers is to charge as quickly as possible. The willingness might be low.	Largest contributor to peak demand Load profiles vary by shift schedule and fleet duty cycles
AC level 2 chargers (auxiliary)	6.6–19.2 kW per port	Slow/overnight charging or topping off	Moderate Predictable daily patterns but varies by user behavior.	Very high	Often used when grid constraints exist Highly schedulable and controllable

Notes: DCFC = direct current fast charger; EV = electric vehicle; HDV = heavy-duty vehicle; MCS = Megawatt Charging System.

Source: Energy Systems Integration Group.

TABLE A-11

Hydrogen Refueling Stations (Non-Electrolyzer)

Component	Power-Consuming Devices	Function / Service	Uncertainty Level	Flexibility Potential	Remarks
Hydrogen compression and dispensing	Mechanical and electrical compressors Cryogenic H_2 pumps High-pressure nozzles (350–700 bar)	Compresses and dispenses H_2 for fuel cell electric vehicles	High Demand driven by fleet/refueling cycles, which may be irregular.	Moderate With buffer tanks, some load shifting possible	Buffer storage decouples electricity usage from vehicle demand
Cooling and safety systems	Pre-coolers Sensors, alarms, venting fans	Prepares H_2 for safe delivery	Moderate Predictable daily patterns but varies by user behavior.	Moderate	Can cycle without disrupting refueling Often backed with battery or thermal storage

Source: Energy Systems Integration Group.

References

Almubarak, S., H. Ibrahim, D. Singhania, and P. Enjeti. 2023. "Energy Consumption and Power Quality in Bitcoin Mining Facilities in Texas." *2023 IEEE Energy Conversion Congress and Exposition.* https://doi.org/10.1109/ECCE53617.2023.10362562.

Barroso, L. A., U. Hölzle, and P. Ranganathan. 2022. *The Datacenter as a Computer: Designing Warehouse-Scale Machines,* 3rd ed. Springer Cham. https://link.springer.com/book/10.1007/978-3-031-01761-2.

Bergmann, D., and C. Stryker. n.d. "What Is Backpropagation?" *IBM Think.* Accessed February 27, 2026. https://www.ibm.com/think/topics/backpropagation.

Choukse, E., B. Warrier, S. Heath, L. Belmont, A. Zhao, H. A. Khan, et al. 2025. "Power Stabilization for AI Training Datacenters." Preprint. arXiv:2508.14318. https://doi.org/10.48550/arXiv.2508.14318.

Colangelo, P., A. K. Coskun, J. Megrue, C. Roberts, S. Sengupta, V. Sivaram, et al. 2025. "Turning AI Data Centers into Grid-Interactive Assets: Results from a Field Demonstration in Phoenix, Arizona." Preprint. arXiv:2507.00909. https://doi.org/10.48550/arXiv.2507.00909.

Delta Electronics. 2025. "Delta Presents Comprehensive Solutions for AI Data Center with Containerized Data Center and HVDC Power Solution at COMPUTEX 2025." https://www.deltaww.com/en-us/news/39729.

Donnellan, D., A. Lawrence, D. Bizo, P. Judge, J. O'Brien, J. Davis, M. Smolaks, J. Williams-George, and R. Weinschenk. 2024. *Uptime Institute Global Data Center Survey 2024.* Uptime Institute. https://datacenter.uptimeinstitute.com/rs/711-RIA-145/images/2024.GlobalDataCenterSurvey.Report.pdf?version=0.

Dynamic Air. 2026. "ECM Fan Array—vs. 3 Phase Motor and VFD." https://dac-hvac.com/ecm-fan-array/.

Eaton. 2025. "XLHV Supercapacitor: High Voltage Module." Data sheet. https://www.eaton.com/content/dam/eaton/products/electronic-components/resources/data-sheet/eaton-xlhv-supercapacitor-module-data-sheet-elx1352-en.pdf.

Eichman, J., K. Harrison, and M. Peters. 2014. *Novel Electrolyzer Applications: Providing More Than Just Hydrogen.* National Renewable Energy Laboratory. https://doi.org/10.2172/1159377.

Fischer, U. R., H.-J. Krautz, M. Wenske, D. Tannert, P. Krüger, and C. Ziems. 2016. "Hydrogen Hybrid Power Plant in Prenzlau, Brandenburg." In *Hydrogen Science and Engineering: Materials, Processes, Systems, and Technology,* edited by D. Stolten and B. Emonts. Wiley-VCH. https://onlinelibrary.wiley.com/doi/10.1002/9783527674268.ch44.

Gan, H., and P. Ranganathan. 2025. "Balance of Power: A Full-Stack Approach to Power and Thermal Fluctuations in ML Infrastructure." *Google Cloud Blog,* February 11, 2025. https://cloud.google.com/blog/topics/systems/mitigating-power-and-thermal-fluctuations-in-ml-infrastructure.

Google. 2025. "2025 PUE Yearly Report." Accessed February 27, 2026. https://datacenters.google/efficiency/.

Google. 2023. *2023 Environmental Report.* https://www.gstatic.com/gumdrop/sustainability/google-2023-environmental-report.pdf.

Heron Power. 2025. "Heron Link: Unlocking Efficient Power Conversion for Gigascale AI." White paper. October 2025. https://heronpower.com/resources/unlocking-efficient-power-conversion-for-gigascale-ai.

Hurley, M. 2020. "Design Considerations for the Deployment of UPS Systems in Data Centers ." *Schneider Electric Blog,* July 20, 2020. https://blog.se.com/datacenter/2020/07/20/design-considerations-deployment-ups-systems-in-data-centers/.

IEA (International Energy Agency). 2025. *Energy and AI.* https://www.iea.org/reports/energy-and-ai.

IEA (International Energy Agency). 2024. *Global Hydrogen Review 2024*. https://www.iea.org/reports/global-hydrogen-review-2024.

IEA (International Energy Agency). 2023. *Global Hydrogen Review 2023*. https://www.iea.org/reports/global-hydrogen-review-2023/executive-summary.

Iyengar, M., and A. Huffman. 2025. "AI Infrastructure Is Hot. New Power Distribution and Liquid Cooling Infrastructure Can Help." *Google Cloud Blog,* April 29, 2025. https://cloud.google.com/blog/topics/systems/enabling-1-mw-it-racks-and-liquid-cooling-at-ocp-emea-summit.

Jain, R., K. Nagasawa, S. Veda, and S. Sprik. 2023. "Grid Ancillary Services Using Electrolyzer-Based Power-to-Gas Systems with Increasing Renewable Penetration." *e-Prime - Advances in Electrical Engineering, Electronics and Energy* 6: 100308. https://doi.org/10.1016/j.prime.2023.100308.

Li, X., K. Ravikumar, and C. Peabody. 2024b. "+/- 400 Vdc Rack for AI/ML Applications." Slide deck from presentation at Open Compute Project Global Summit, San Jose, CA, October 15-17, 2024. https://drive.google.com/file/d/1BCiJ5XJg1bbhF7l_BKvnSzjoUClUdGk2/view.

Li, Y., and Y. Li. 2025. "AI Load Dynamics–A Power Electronics Perspective." Preprint. arXiv:2502.01647. https://doi.org/10.48550/arXiv.2502.01647.

Mehra, V., and R. Hasegawa. 2023. "Supporting Power Grids with Demand Response at Google Data Centers." *Google Cloud Blog,* October 3, 2023. https://cloud.google.com/blog/products/infrastructure/using-demand-response-to-reduce-data-center-power-consumption.

Mitra, P. 2025. "Data Center Modeling for Transmission Planning Studies." Slide deck. EPRI. https://www.ercot.com/files/docs/2025/07/10/EPRI_Data-Center-Modeling-for-Transmission-Planning-Studies_ERCOTLLWG.pptx.

Morris, S. 2025. “Data Center Inside ERCOT: Observations and Patterns.” Slide deck from presentation at ERCOT Large Load Working Group meeting, August 14, 2025. https://www.ercot.com/files/docs/2025/08/11/LLWG_DataCenterObs.pptx.

NERC (North American Electric Reliability Corporation). 2025a. “Characteristics and Risks of Emerging Large Loads.” Large Loads Task Force White Paper. https://www.nerc.com/globalassets/who-we-are/standing-committees/rstc/whitepaper-characteristics-and-risks-of-emerging-large-loads.pdf.

NERC (North American Electric Reliability Corporation). 2025b. “Large Loads Task Force Meeting.” October 23, 2025. https://www.nerc.com/globalassets/who-we-are/standing-committees/rstc/lltf/lltf_presentations_october_23_2025.pdf.

Nguyen, B. L. H., M. Panwar, R. Hovsapian, Y. Agalgaonkar, and T. Vu. 2022. “Hierarchical Control of Grid-Connected Hydrogen Electrolyzer Providing Grid Services.” *2022 IEEE 1st Industrial Electronics Society Annual On-Line Conference (ONCON)*. https://doi.org/10.1109/ONCON56984.2022.10126806.

NLR (National Laboratory of the Rockies). 2025. “Renewable Electrolysis.” https://www.nlr.gov/hydrogen/renewable-electrolysis.

NVIDIA. 2025. “NVIDIA Blackwell Ultra DGX SuperPOD Delivers Out-of-the-Box AI Supercomputer for Enterprises to Build AI Factories.” Press release, March 18, 2025. https://nvidianews.nvidia.com/news/blackwell-ultra-dgx-superpod-supercomputer-ai-factories.

OpenAI. 2018. “AI and Compute.” OpenAI blog, May 16, 2018. https://openai.com/index/ai-and-compute/.

Patel, P., E. Choukse, C. Zhang, Í. Goiri, B. Warrier, N. Mahalingam, and R. Bianchini. 2024. “Characterizing Power Management Opportunities for LLMs in the Cloud.” *Proceedings of the 29th ACM International Conference on Architectural Support for Programming Languages and Operating Systems, Volume 3.* https://doi.org/10.1145/3620666.3651329.

Patterson, D., J. Gonzalez, Q. Le, C. Liang, L. M. Munguia, D. Rothchild, D. So, M. Texier, and J. Dean. 2021. “Carbon Emissions and Large Neural Network Training.” Preprint, arXiv:2104.10350. https://doi.org/10.48550/arXiv.2104.10350.

Ross, B. 2025. “UPS Fault Dynamics.” Slide deck from presentation at Electric Reliability Council of Texas Large Load Working Group meeting, July 11, 2025. https://www.ercot.com/files/docs/2025/07/10/PNNL_ERCOT_UPS_Presentation_Outline.pptx.

Salim, M. 2026. “Data Center Facility Infrastructure Trends to Support HPC and AI.” *7x24 Exchange Magazine.* https://www.7x24exchange.org/industry-news/magazine-archives/data-center-facility-infrastructure-trends-to-support-hpc-and-ai/.

Samanta, A., S. Majumder, H. Ibrahim, P. Enjeti, and L. Xie. 2024. “Electromagnetic Transient Model of Cryptocurrency Mining Loads for Low-Voltage Ride Through Assessment in Transmission Grids.” *2024 IEEE Power and Energy Society General Meeting (PESGM).* https://doi.org/10.1109/PESGM51994.2024.10689224.

Sapre, S., and Z. Sras. 2025. "OCP Presentation 18kW DC PSU 100kW DC Power shelf," presentation, July 9, 2025.

Schwartz, R., J. Dodge, N. A. Smith, and O. Etzioni. 2020. "Green AI." *Communications of the ACM* 63(12): 54–63. https://doi.org/10.1145/3381831.

Siemens Energy. 2026. "SVC PLUS FS® (E-STATCOM)." https://www.siemens-energy.com/global/en/home/products-services/product/svcplus-frequency-stabilizer.html.

Srivastava, N., and H. Petty. 2025. "Building the 800 VDC Ecosystem for Efficient, Scalable AI Factories." *NVIDIA Technical Blog,* October 13, 2025. https://developer.nvidia.com/blog/building-the-800-vdc-ecosystem-for-efficient-scalable-ai-factories/.

Strubell, E., A. Ganesh, and A. McCallum. 2019. "Energy and Policy Considerations for Deep Learning in NLP." *Proceedings of the 57th Annual Meeting of the Association for Computational Linguistics*, Florence, Italy, July 28 - August 2, 2019, pp. 3645–3650. https://www.aclweb.org/anthology/P19-1355.pdf.

Tavanaee, M., J. Girona-Badia, M. Cheah-Mañe, O. Gomis-Bellmunt, and E. Prieto-Araujo. 2025. "Grid-Connected Power Converters for Electrolysis Applications." *IEEE Access* 13: 187211–187225. https://doi.org/10.1109/ACCESS.2025.3627384.

Tesla. 2025. "Battery Storage Applications at Data Centers." Slide deck. Presentation at the Electric Reliability Council of Texas Large Load Working Group, May 2025. https://www.ercot.com/files/docs/2025/05/15/11-Megapack-at-Data-Centers-_-ERCOT-LLWG_5.16.25_vShare.pdf.

Thompson, N. C., K. Greenewald, K. Lee, and G. F. Manso. 2022. "The Computational Limits of Deep Learning." Preprint, arXiv:2007.05558.https://doi.org/10.48550/arXiv.2007.05558.

Wheeler, K. A., A. W. Bowers, C. H. Wong, J. Y. Palmer, and X. Wang. 2018. "A Power Quality and Load Analysis of a Cryptocurrency Mine." 2018 IEEE Electrical Power and Energy Conference*,* Toronto, ON, Canada, 2018, pp. 1-6, https://doi.org/10.1109/EPEC.2018.8598358.

Zeng, Y., Y. Qiu, L. Xu, C. Gu, Y. Zhou, J. Li, S. Chen, and B. Zhou. 2026. "Optimal Investment Portfolio of Thyristor- and IGBT-Based Electrolysis Rectifiers in Utility-Scale Renewable P2H Systems." *IEEE Transactions on Sustainable Energy*, 17(2): 740–755. https://doi.org/10.1109/TSTE.2025.3595150.

Large Loads: Behaviors, Capabilities, and Limitations

A Report by the Energy Systems Integration Group's Large Loads Task Force

This report is available at https://www.esig.energy/resources/reports-and-briefs/.

To learn more about the ESIG Large Loads Task Force and the recommendations in this report, please visit https://www.esig.energy/working-groups/large-loads/ or send an email to info@esig.energy.

The Energy Systems Integration Group is a nonprofit organization that marshals the expertise of the electricity industry's technical community to support grid transformation and energy systems integration and operation. More information is available at https://www.esig.energy.